Essays in Applied Microeconometrics

Dissertation submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Johannes Sebastian Matthias Nikolaus Kunz
from Germany

approved in July 2016 at the request of

Prof. Dr. Rainer Winkelmann
Prof. Dr. Gregory Crawford

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zürich, July 20^{th} 2016

Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

Acknowledgments

Beyond acknowledging comments and suggestions on the individual chapters, which are given therein, I would like to thank a few people that supported me during the process of writing this dissertation. First and foremost, I am indebted to Rainer Winkelmann, for whom I worked as a research assistant at the Chair of Statistics and Empirical Economic Research during the completion of this dissertation. This dissertation benefited enormously from his critical questioning and remarks on each of its chapters. He patiently answered a myriad of my questions which helped me to gain insights into econometrics far beyond what I could have hoped to achieve. I cannot thank him enough for his encouragement and support throughout this time; I also want to thank Gregory Crawford for co-advising this thesis and for his interest in my work.

I would further like to thank Kevin Staub, Gregori Baetschmann, Florian Schaffner, Raphael Studer, and Steven Stillman for sharing their insights into econometrics; they have been a constant source of inspiration. Countless discussions and Skype-calls with Kevin Staub have been among the greatest pleasures during the time; I also want to thank him for inviting me to the University of Melbourne. Furthermore, I highly appreciated the academic and even more so the personal support of Sabrina Studer and Philippe Ruh. I admire each of these fellow colleagues, with whom I became dear friends and I look forward to many more interactions.

I am also grateful to Uschi Backes-Gellner and the Swiss Leading House Economics of Education, which granted the third chapter of this dissertation the Best Paper Runner-Up Award 2014. I appreciated the opportunities given to me to present my work; among the most valuable experiences were the conversations with Ed Lazear which always posed an intellectual challenge. I also appreciated various philosophical discussions with Josef Falkinger.

Finally, I thank Katherina Giese for her encouragement, and much more so for her constant and loving support throughout this time.

Contents

List of Figures

List of Tables

Dissertation overview

The three chapters of this dissertation contribute to a diverse set of current topics in labor economics and, by introducing new estimation techniques, to the applied microeconometrics literature. Each chapter makes a methodological as well as a substantive contribution. In this overview, I first summarize the chapters' individual contributions, before highlighting their common grounds.[1]

Chapter 1 titled *Subjective completion beliefs and the demand for post-secondary education* is written jointly with Kevin Staub. In this chapter, we study the role of uncertainty in individuals' post-secondary education choice. Although investment in human capital is a classical topic in labor economics, evidence on the impact of uncertainty is still relatively scarce. Early studies introducing uncertainty relied heavily on structural assumptions; most commonly on rational expectations (that is individuals make on average correct predictions). These and other less restrictive assumptions have been called into question by several recent studies, which show elicited subjective beliefs to be superior to assumptions-based imputations when assessing individual's education choices (i.e. Huntington-Klein, 2015*b*; Stinebrickner and Stinebrickner, 2012; Zafar, 2011*a*).

Most existing research in this area focuses on uncertainty about earnings or employment prospects. At this point the verdict is still out, whether underinvestment results from misaligned subjective beliefs about objective labor market information. However, most convincing evidence

[1]Full citations details can be found in the references of the respective chapters.

based on randomized information-experiments suggests that it is not the lack of labor market information that governs individual's choices (i.e. Kerr et al., 2014; Fryer, 2013). In contrast, we focus on the role of uncertainty in completing an educational degree. Completion uncertainty is a broader concept as it entails not only pecuniary but also non-pecuniary returns whose relevance to educational choice have been confirmed in several studies. A recent survey is given by Oreopoulos and Salvanes (2011). Similar to pecuniary returns, non-pecuniary returns are likely to depend on graduation. In the chapter, we discuss additional reasons why subjective completion beliefs might be more important to an individual's choice than subjective earning or employment beliefs.

Turning to our results, we first establish that subjective beliefs about finishing an educational track are important determinants of post-secondary education choices and that they have long-lasting impacts on the adolescents' educational careers: they are highly predictive of aspirations (measured concurrently with the beliefs), actual investments (measured at least two years later), and actual completion (measured at least five years later). By varying specifications, including a large set of control variables, and bounding the effect against potential selection-on-unobservables we show that this is a very robust finding.

Most available evidence using subjective beliefs (about the state of the labor market) to study uncertainty in educational choices extends only to college-major choices of students already enrolled in a specific institution. Broadly speaking, these studies find that early subjective beliefs (at college enrollment) continue to be important for educational choices and outcomes later on (Zafar, 2011*b*) and are revised in response to information revealed by academic performance (Stinebrickner and Stinebrickner, 2014*b*). Exploiting a large-scale population survey in Germany, we extend the literature by presenting evidence that the formation of subjective beliefs takes place already while in secondary education. Thus, they date further back than previously found. Moreover, subjective completion beliefs are relevant in the overall population of adolescents and not specific to college major choices (of students already enrolled in a specific institution). In addition, using a rich set of covariates, we find that both subjective beliefs and educational aspirations are strongly related to academic ability and personality traits. By contrast, actual enrollment and completion depend to a larger extent on family characteristics, the state of the local labor market, and the regional supply and demand in the post-secondary education market. This suggests a potential for informational policy interventions.

Despite the advantages of our dataset, it is not without caveats: the subjective beliefs are not repeatedly assessed and are not elicited separately for counterfactual education choices. We further disaggregate the impacts of subjective beliefs into the choices between educational tracks: apprenticeship, high school and apprenticeship, and university. Due to the lack of counterfactual beliefs, we condition on the students' educational aspirations when disaggregating to educational tracks. In confirmation with the existing literature, we find GPA to be the main driver also of subjective completion beliefs for students choosing an university education. Yet, subjective beliefs of those choosing a vocational education seem to be driven by other characteristics. For those students, the subjective beliefs remain a substantial predictor for future investments after accounting for GPA on top of a large set of individuals' characteristics.

In the last part of the chapter, we develop and estimate a structural education choice model that accounts for unobserved preferences for post-secondary education, forward-looking behavior, and the sequentiality of choices. Accounting for the option value that arises from the continuation possibility after finishing high school, our results reveal that the subjective completion beliefs are essential to the choice of an apprenticeship (as before), but also to the choice of obtaining a high school degree (which confirms and extends the results found by Pinger, 2015). After finishing high school, they are irrelevant to the choice between an apprenticeship training and university studies. This might be explained by updating in response to information revealed by grades obtained in high school. Moreover, we find that the subjective completion beliefs to be most decisive for adolescents with low academic ability and weak preferences for education; a group of high policy relevance that has largely been ignored in the present literature. As more and more countries striving to introduce an apprenticeship system (Obama, 2014, State of the Union Address), this finding is of first-order relevance to policy-makers.

Methodologically, the chapter makes two simple but seminal contributions that are likely to be important for future research in this area. First, we extend the framework to bound selection-on-unobservables of Altonji, Elder and Taber (2005*a*,*b*, 2008) to a non-binary endogenous random variable. Second, we extend the structural education choice model of Taber (2000, 2001) to the German educational system, a system with multiple education streams.

Chapter 2, titled *An econometric model of health care demand with non-linear pricing* is written jointly with Rainer Winkelmann and forthcoming in *Health Economics*. In the chapter, we introduce a new structural estimation model to study health care demand (as measured by doctoral visits). Our model is well suited to study a new payment scheme introduced in Germany: Between 2004 and 2012, patients had to pay a one time fee for the first-visit in a calendar quarter, all subsequent visits in that quarter were free of charge.

The decision to visit a doctor is subject to an extensive literature in health economics. A stylized finding is that many people never visit a doctor. However, once they visit a doctor their likelihood of further visits increases (i.e. Pohlmeier and Ulrich, 1995). This led to the conclusion that the first visit is structurally different from subsequent visits, justifying a policy targeting the first visit. A class of models accounting for this difference is hurdle models, which we extend by allowing the timing of the first choice to influence subsequent choices. This naturally fits situations in which the first visit influences the costs of subsequent visits. We further extend the model to a panel framework, which allows us to account for unobserved heterogeneity and show that our model can easily be adapted to reporting mismatches (when the reporting period overlaps two payment periods). In the appendix to the chapter we present estimation details and Monte Carlo simulations. The simulations show that the model works well in small samples.

Turning to the policy intervention, by changing the cost of visiting a doctor the policy was explicitly designed to reduce the number of doctoral visits. To identify the reform effect we use a difference-in-differences strategy, where the privately-insured (which were not subject to the reform) serve as a counterfactual for the publicly insured individuals. Previous research using the same identification and policy change found mixed results of the policy. Augurzky, Bauer and Schaffner (2006) and Schreyögg and Grabka (2010) find no (significant) effect, whereas Farbmacher and Winter (2013) find a significant negative effects of the policy upon the number of doctoral visits. Farbmacher and Winter (2013) suggest that the different results are due to the fact that for most individuals the reporting period was different from the payment period. However, their introduction of the reporting mismatch is rather ad hoc. By contrast, it can be introduced very naturally in our model. In sum, across various estimation models that account for unobserved heterogeneity and reporting mismatch, we find no evidence that the policy had a significant effect upon the number of doctoral visits.

Chapter 3, titled *Analyzing educational achievement differences between second-generation immigrants: Comparing Germany and German-speaking Switzerland* is published in the German Economic Review. In this chapter, I promote a new approach to assess immigrants' children learning achievements by comparing them across countries.

Ideally, when assessing immigrant children one would like to use standardized achievement test data which contains information on the policy area, migration history of the individual, and a sufficient sample size to assess immigrants' children in detail. In most European countries neither is available; most importantly, the only large scale standardized achievement test such as PISA, PRILS, or TIMSS do not allow for regional assessments of students across educational institutions (i.e. federal states in Germany). The dominant approach in the literature to learn about institutional differences is to compare the immigrant children to their native peers and to compare their test score differences across countries (i.e. Dustmann, Frattini and Lanzara, 2012). This approach is suited to assess inequality within the educational systems. However, it is flawed for the assessment of learning achievements by immigrant children. This is due to the fact that native pupils are different in many aspects and, as I argue, are an inappropriate counterfactual for the second-generation immigrants' achievements. It is, for example, typical in the literature to assess immigrant-native achievement gaps conditional on language spoken at home. Although the natives who do not speak the national language at home are either not existent (hence predicted) or most likely a very different group of students (i.e. third-generation immigrants). A related problem is that in these large scale student assessments there is no information on pre-migration characteristics, such as the reason for or the time of migration, which would, at least partly, help to control for potential self-selection of first-generation migrants to host countries.

To advance our understanding, I propose a new approach to study second-generation immigrant students by comparing them directly across countries. Focusing exclusively on second-generation –as opposed to first-generation– immigrants has the advantage that they are born in the country of testing and shared the entire school system, which limits the impacts of the unobserved pre-migration environment. Obviously, comparing immigrants or their children across countries creates another problem. Immigrants self-select into countries (even more than into locations/policy areas), hence caution is warranted when applying this approach. I therefore compare Germany to German-speaking Switzerland which, as a I argue, are well suited for such a comparison as both regions experienced a very similar migration history. This aligned reasons for migration and countries of origin of the first-generation immigrants. As a side effect, their countries of origin are sufficiently overlapping, which allows to compare students from the same home country in different host countries. The lack of overlap in migrants ancestry across host countries is a recurring problem in the existing literature. Moreover, as both regions have the same testing language (German), the reading test scores can validly be compared, which is paramount to immigrants' learning achievement and integration into the host society. Despite these advantages, the comparison of Germany to German-speaking Switzerland is of course not without problems, as it might still be the case that the immigrants in Germany are different from those in Switzerland. I therefore control for a substantial set of controls and use a matching decomposition technique to assure that only comparable students are compared.

On the substantive side, I show that children of immigrants in Switzerland are performing much better than their counterparts in Germany and these differences cannot be explained by observable background characteristics. By decomposing this effect along the test score distribution, I find that the differences mainly stem from very low performing children of immigrants. The most crucial difference seems to be the language spoken at home. When it is different from German, it always increases the gap that cannot be explained by the students' background characteristics. These differences are robust to the inclusion of the parents' country of origin. Additionally, Switzerland seems to be particularly beneficial for unfavorably-endowed children of immigrants and children of Turkish descent, while being relatively less beneficial for children of native-born parents. The almost equal performance of natives suggest that there is an integration-specific reason for the enhanced performance in German-speaking Switzerland. Any analysis of institutional differences is necessarily suggestive, however, my results point to a role of segregating students into classes of low-performers in Germany as compared to German-speaking Switzerland, which almost fully accounts for the differential performance.

To sum up, both chapters 1 and 3 contribute to the economics of education, and have important implications for disadvantaged students; chapter 2 contributes to the health economics literature, and how to better assess policy interventions targeting the number of doctoral visits. Although the three chapter presented in this dissertation cover a diverse set of contents, all advance current topics in labor economics by introducing new microeconometric estimation techniques to their respective subfields. Chapters 1 and 2 develop and extend dynamic estimation models of individuals' decision-making. In these, choices taken by the individual impact their future decision-making. In addition to this dynamic aspects, both empirical models account for unobserved heterogeneity. In both chapters advanced maximum likelihood procedures are developed, programmed, and will be made publicly available. In contrast to the parametric approaches used in chapters 1 and 2, chapter 3 introduces semi-parametric matching technique and a sound approach to study immigrant students' learning process.

This dissertation shows how applied research can benefit from careful implementation of microeconometric estimation techniques. Currently, there is an apparent skepticism in mainstream economic literature against thoughtful modeling of economic behavior (although very recently the pendulum appears to be swinging back again). The change in the empirical paradigm was let by the desire to establish credible causal claims, which can only be assured with exogenous variation (based on experiments). From the perspective of this dissertation, there will always be a back and forth between more or less theory-dependent approaches. Establishing causality with as few as possible assumptions is an indisputable objective of all social sciences. However, results of experimental and quasi-experimental variation are necessarily application- and circumstance-specific. Thus, to take credible findings to greater use, it is necessary to put these back into models to extend their implications and reach. Dogmatist proponents of either side are important for the advancement of science, but so are undogmatic researchers making the tools from both sides available and easily accessible to researchers and policy-makers with an applied interest.

Based on these arguments, the models and approaches pursued in this dissertation are aimed to add as much structure as necessary and as little as possible. On the one side reducing or relaxing the assumptions in structural models and on the other advancing the treatment effect methodology to more model based implications. This reasoning is the basis of introducing assumption-free subjective beliefs in a very simple two-stage decision process, in chapter 1. Or, to identify a structural model based on a natural experiment, i.e. a group unaffected by the policy used as a counterfactual, as in chapter 2. Or, to interpret a classical topic of inequality in a treatment effect perspective which implies that natives cannot serve as an adequate control for immigrant children, as in chapter 3. Introducing simple structure in combination with experimental or quasi-experimental variation allows researchers to draw more useful conclusions than from exogenous-variation-assumptions or adding-as-much-structure-as-possible alone. Advances presented in the following chapters based on the idea of combining credible identification with implications from behavioral models include bounding effects against unobservables (chapter 1), assessing unobserved preferences for education (chapter 1) or unobserved heterogeneity in doctoral visits (chapter 2), and comparing immigrants across host countries (chapter 3).

Chapter 1

Subjective completion beliefs and the demand for post-secondary education

This chapter is jointly written with Kevin E. Staub.[1]

The outcome of pursuing a post-secondary educational degree is uncertain. A student might not complete a chosen degree for a number of reasons, such as academic insufficiency or financial constraints. Thus, when considering whether to invest in post-secondary education, students must factor in their completion probability into their decision. We study the role of this uncertainty in educational choices using students' subjective beliefs about completing a post-secondary education, which were elicited prior to students' completing secondary education. We relate these subjective completion probabilities to their subsequent educational choices and outcomes using representative survey data from Germany. Following the students over time, we find that the initial beliefs are predictive of intentions to invest in education, actual subsequent educational investments, and degree completion. We assess the heterogeneity of the impact across different educational paths. After controlling for academic ability, we find that subjective beliefs are most relevant in choosing a vocational education. In addition to reduced form models, we estimate a structural choice model of sequential investment in education that allows for unobserved tastes and preferences for education and forward-looking behavior. The results confirm the influence of subjective completion beliefs on choosing a post-secondary education.

Keywords: Subjective beliefs, Educational completion uncertainty, Human Capital Investment.
JEL classification: I21, I26, J24;

[1]We thank Joseph Altonji, Orazio Attanasio, Uschi Backes-Gellner, Dan Black, Timo Boppart, Gregory Crawford, Thomas Dohmen, Ernst Fehr, Joe Hirschberg, Jamie Gloor, Uli Kaiser, Ed Lazear, Edwin Leuven, Jenny Lye, Philippe Ruh, Florian Schaffner, Carmit Segal, Andreas Steinhauer, Steven Stillman, Sabrina Studer, Rainer Winkelmann, Joachim Winter, and seminar participants at Kloster Kappel, Luzern, Melbourne, Schaffhausen, Zurich, the Ski and Labor Seminar, Laax, and the Workshop of Personnel Economics and Economics of Education, Lech, for helpful comments and suggestions, and Daniel Auer and Luca Tonizzo for very helpful research assistance.

1.1 Introduction

Educational choice is one of the most important career decisions young adults have to make, and one that must be made under partial information. Students not only face the difficulty of having to predict labor market prospects for several educational degrees —an endeavoring task even for highly distinguished scholars (Manski, 1993)— but also the challenge of having to predict their own returns to each of these degrees. When choosing an educational track, students further need to foresee their own abilities for that new educational environment and their chances of succeeding in such an environment. Understanding the role of uncertainty in individuals' post-secondary educational choices is essential for designing effective educational policies. For instance, if students' expectations are misaligned, providing additional information can be a cost-effective measure to enhance educational choices, and eventual career success.

In this study, we assess the role of uncertainty about completing an educational degree in young adults' education choices. We show how subjective beliefs about completing a post-secondary education, elicited while in secondary education are important predictors of post-secondary educational aspirations, enrollment, and completion. We find that both beliefs and educational aspirations are strongly related to academic ability and personality traits. Conversely, actual enrollment and completion depend to a larger extent on family characteristics, the state of the local labor market, and the regional supply and demand in the post-secondary education market, which suggests a potential for informational policy interventions. In addition, we assess effect heterogeneity across post-secondary educational tracks. Academic ability appears to be the main driver of subjective beliefs in choosing a university education. In contrast, subjective beliefs of those choosing a vocational education seem to be driven by other characteristics. Finally, we develop and estimate a structural educational choice model that accounts for unobserved preferences for post-secondary education, forward-looking behavior, and the sequentiality of choices. We find that the subjective completion beliefs are most decisive for adolescents with low academic ability and weak preferences for education; a group that has largely been ignored in the present literature.

The context of this study is the secondary and post-secondary educational system in Germany, which is renowned for its well-functioning apprenticeship system. Apprenticeship systems are now tested and implemented in several countries (including the US, cf. President Barack Obama's State of the Union Address, 2014) motivated by the low youth unemployment rates observed in countries with apprenticeship systems. In Germany, when finishing secondary education at the age of 16 to 17, young adults choose between dropping out of education, investing in an apprenticeship, or continuing with general education that enables them to enroll in an university. In addition to analyzing the demand for education under conditions of uncertain outcomes, we study how the demand varies across educational tracks, which is relevant to policy makers who aim to introduce apprenticeship systems.

In general, when studying choice under uncertainty, researchers have to assume how expectations are formed (Manski, 2004). Most commonly, researchers impose rational expectations; e.g., that individuals' predictions, usually about future wage distributions, are unbiased. In the context of individuals' educational choices it is important to note that even if students were able to accurately predict the future wage distribution, their (perceived) internal rates of return —the rates upon which they act— might be very different from the aggregate returns.[2] By far, the most widely used alternative is to use direct measures of elicited subjective beliefs, which circumvents these problems (Manski, 2004). Several studies show that the use of elicited expectation data can be superior to those constructed using rational expectation assumptions, and they are meaningful measures in educational choice models (e.g., Attanasio and Kaufmann, 2014; Huntington-Klein, 2015*b*; Stinebrickner and Stinebrickner, 2012; Zafar, 2011*a*). Although the literature on educational decision making under uncertainty using elicited subjective beliefs is rapidly growing, we advance the literature in important dimensions. We assess the role of prior subjective beliefs formed in secondary education in a representative population survey and follow these adolescents over time until they complete their post-secondary education.[3] Second, much of the existing literature on the demand for post-secondary education focuses

[2]Several approaches were proposed to circumvent rational expectations. Early approaches based on structural assumptions that distinguish *ex ante* from *ex post* returns include Carneiro et al. (2003), Cunha, Heckman and Navarro (2005), and Cunha and Heckman (2007). Their framework is also applied recently in Foley, Gallipoli and Green (2014). Another approach is to include measures of uncertainty within the expected wage functions of Roy-type selection models; for example, Mazza (2014) introduces the (rational expectation) variance of earnings and Fossen and Glocker (2014) include risk preferences.

[3]Also related to our study is the evolving literature on college major choice using subjective beliefs: Arcidiacono, Hotz and Kang (2013), Arcidiacono et al. (2014), Hastings et al. (2015), Huntington-Klein (2015*c*), Stinebrickner and Stinebrickner (2014*a*), and Wiswall and Zafar (2015*a*). In Germany there are no majors, as students specialize at the beginning of their studies. However, we follow a similar approach as these studies by allowing for selection into different educational tracks.

on investment, rather than on aspirations or completion.[4] Thus, we assess each of these three outcomes while also accounting explicitly for uncertainty in students' choices.

In contrast to the existing literature that investigates the uncertainty about wages or the likelihood of unemployment, our main focus is on completion uncertainty. Although some theoretical work includes completion uncertainty (e.g., Comay, Melnik and Pollatschek, 1973; Manski, 1989; Altonji, 1993), there is little empirical work in this area. Theoretical studies emphasize the sequentiality of the educational decisions and that "[d]ifferences in dropout probabilities may be more important than differences in ex post payoffs in determining the ex ante return to attending a particular school," (Altonji, 1993, p74).[5] This hypothesis is empirically supported by Hussey and Swinton (2011), based on a predicted likelihood of completion. However, such predicted completion probabilities are limited in that they are only a crude proxy for the subjective beliefs on which people act. We contribute to this literature by integrating elicited subjective completion probabilities into a sequential model of educational choice. In this respect, our analysis is most closely related to Wiswall and Zafar (2015*a*), which also uses students' subjective completion beliefs. Our research addresses complementary questions such as how the choice process differs for adolescents not enrolled in college and how these beliefs relate to actual completion. Our paper is the first to study subjective completion beliefs assessed before the end of secondary education in a population survey in the context of a detailed educational investment model.

One way in which completion uncertainty affects educational choice is by simply amplifying *ex ante* wage uncertainty. However, completion uncertainty may have important consequences beyond that general channel. For example, various non-pecuniary aspects have been shown to be relevant to educational choice (see Oreopoulos and Salvanes, 2011, for a recent summary). In order to benefit from them, staying in the chosen educational path and/or completing the degree might be crucial. For instance, studies using elicited subjective beliefs about labor market prospects consistently find the (non-financial) consumption value of education or major-

[4]One reason is that the data on completion is necessarily incomplete: individuals can always come back and acquire more education. For a detailed discussion of educational completion, see Turner (2004) and Bound and Turner (2011). Notable exceptions are Venti and Wise (1983) and Light and Strayer (2000). Similarly, the literature on aspirations is still comparatively small, although it has been growing recently (e.g., Christofides et al., 2015; Wiswall and Zafar, 2015*b*; Zachary and Zafar, 2015).

[5]Manski (1989) raises an important point by clarifying that drop-out rates are not necessarily undesirable from a social planner's point of view: since educational outcomes are uncertain, schooling should be evaluated based on *ex ante* returns rather than on *ex post* success rates.

specific unobserved tastes to be the main drivers of educational choices (i.e., Huntington-Klein, 2015*a*; Wiswall and Zafar, 2015*a*).[6] Such preference-related factors are not affected by pure labor market uncertainty, but they can be affected by completion uncertainty.[7] Our results also point to unobserved preferences for a post-secondary education that play a substantial role in students' choices.

Our study is also closely related to the literature on learning about one's own academic ability (or preferences).[8] The central finding in this literature is that learning about one's own ability is based mainly on academic ability conveyed by students' grade point averages [GPA] (e.g., Stinebrickner and Stinebrickner, 2012, 2014*b*; Zafar, 2011*b*). Although these studies offer a valuable assessment of the subjective beliefs at various points in time and in great detail, thus far they have focused on single institutions rather than a representative sample. Milla (2014) adds to and supports the generalizability of the previous findings by studying aspiration updating in responses to changes in GPA using a population survey of college students. Still, such a design imposes a sample selection. By exclusively focussing on college students it ignores young adults who dropped out of education because they were less optimistic about their educational prospects. We contribute to this literature by assessing initial subjective beliefs prior to college enrollment in a representative survey population and by providing evidence on both beliefs and educational aspirations at this early stage. Our evidence supports and extends Zafar's presumption that "prior belief[s] [at the start of college] continue[s] to be important. In attempting to understand the choice of college majors, it might be useful to focus on students at earlier stages of their schooling (for example, in high school) and analyze their subjective beliefs" (Zafar, 2011*b*, p339f).[9]

[6]Similar evidence comes from more structural approaches that do not rely on subjective beliefs. For instance, D'Haultfoeuille and Maurel (2013) use a sophisticated Roy model and find non-pecuniary aspects to be predominant in educational choice.

[7]Evidence whether the provision of information about the labor market induces students to invest more in education is mixed, which can be interpreted as broadly in line with our view that there is more to uncertainty than pure wage uncertainty. Supporting evidence comes from developing countries, for instance, see Jensen (2010) for evidence from Dominican Republic and Nguyen (2008) for Madagascar. Oreopoulos and Dunn (2013) find that high school students in Canada update their beliefs in the context of an information experiment. Yet in Finland, Kerr et al. (2014) find that —while students do update their beliefs— there is no significant effect on enrollment; similar results are reported in Fryer (2013). Assessing students' choice process in more detail is therefore highly valuable.

[8]Bulman (2015) shows that providing young adults with better information about their own ability impacts enrollment and college graduation. He finds that important factors other than aptitude deter college attendance, which might be explained by subjective beliefs about educational outcomes.

[9]Due to data limitations, we do not examine subjective beliefs at multiple time points. A detailed analysis of the process behind learning about one's own ability and the evolution of subjective beliefs is beyond the scope of this study, but remain key questions for future research.

Throughout our analyses we account for personality skills, which have been highlighted as main determinants of educational success (see Almlund et al., 2011; Borghans et al., 2006, and references therein). In particular, we show how subjective beliefs relate to the Big Five personality measures, risk attitudes, and locus of control, all of which are now ubiquitous in economic applications (see, for example, Borghans et al., 2006; Dohmen et al., 2010; Caliendo, Cobb-Clark and Uhlendorff, 2015). Of special interest to our design is the locus of control, as Coleman and DeLeire (2003) hypothesize that students with a more internal locus of control (i.e., students who believe their actions affect their outcomes) have higher subjective beliefs about their own returns to education, which increases their efforts and investments in their human capital. Our results support the hypothesis that one's locus of control affects educational choices via subjective beliefs.

Finally, our study is related to recent contributions assessing the role of subjective beliefs as a mediator and a potential explanation of educational differentials in parental unemployment (Pinger, 2015), family background (Keller and Neidhöfer, 2014), or gender and migration (Tolsma, Need and De Jong, 2010). Our framework might prove useful in studying the mediating role of subjective beliefs, since it integrates investment in both secondary and tertiary education jointly in both reduced-form and structural models.

In sum, the main contribution of this study is to provide a better understanding of uncertainty in educational choices and a broad assessment of subjective completion beliefs of young adults. Our analyses include how beliefs are determined and how beliefs relate to intentions to invest in education, actual investments, and degree completion. We explicitly account for the sequentiality of choices and forward-looking behavior of individuals. Moreover, we relate students' beliefs to individual characteristics, family background, personality skills, regional labor and education market conditions, and unobserved tastes and preferences for education. The remainder of this study proceeds as follows: In Section 1.2, we describe the institutional features of the educational system in Germany and present the data we use. In Section 1.3, we assess determinants of subjective completion beliefs. In Section 1.4, we relate the beliefs to educational outcomes and present how the impact of the subjective beliefs varies with selection on observables and unobservables. In Section 1.5, we presents effect heterogeneity across different educational tracks, and in Section 1.6 we develop and estimate a structural model of sequential educational choice. Section 1.7 concludes our paper, briefly summarizing our key findings.

1.2 Institutional setting, data, and descriptive statistics

Institutional setting

A simplified version of Germany's educational system is depicted in Figure 1.1, in which we briefly summarize the system's key features that are relevant to our analysis (more information can be found in Wölfel and Heineck, 2012).

— — — Figure 1.1 about here — — —

The German educational system is characterized by early tracking, which takes place after grade 4 (elementary school), at age 9 to 11 years.[10] Based on grades and teachers' recommendations, the children are tracked into three streams according to their academic ability.[11] The statistical agency in Germany (Statistisches Bundesamt, 2014, p27) reported that in 2012, 10% of children were assigned to the lower track, 19% to the intermediate track, 40% to the upper track (high school), and the remaining children visited other, so-called comprehensive schools that essentially follow the same structure without separating the children.

At the time of entering the survey population, the young adults —ages 16 to 17 years— are in the midst of deciding upon a professional education according to their track. Students completing lower or intermediary tracks have the opportunity to apply for and to start a profession-specific apprenticeship or a vocational education.[12] Although investing directly in an apprenticeship is the dominant path, the young adults can alternatively enroll in a consecutive school-track that leads to the *university entrance qualification* (German: *Abitur*), an equivalent of a high school degree.[13] This high school degree can also be a valuable asset for students who do not want to attend an university. When applying for highly competitive apprenticeship positions, students with a high school degree typically have better chances compared to their

[10]With the exceptions of Berlin and Brandenburg, which track after grade 6 (ages 11 to 13 years).

[11]The binding nature of these recommendations varies across states.

[12]Students who started an apprenticeship before entering the survey population are excluded from our analyses. However, in 2011, only 10.6% started an apprenticeship before the age of 17 years (Statistisches Bundesamt, 2013, p17).

[13]Due to the limited time horizon of our sample we focus on early investment. The possibility of visiting complementary courses that allow students to go to university after apprenticeship completion is not modeled separately. We discuss the implications of this for the interpretation of our results below.

peers who completed a lower track. Some apprenticeship positions are even exclusively available to such students. In 2010, 20.9% of the newly signed apprenticeship contracts went to students holding a high school degree (Statistisches Bundesamt, 2011, p1004). Thus, we model this path separately and refer to it as tertiary apprenticeship.

The decision to start an apprenticeship is somewhat different for students already enrolled in high school. In principle, they can also drop out to start an apprenticeship or continue their high school education and after finishing go on to university or a tertiary apprenticeship.[14] Yet, their default choice is certainly different as they are already enrolled in high school and they do not have to make an active choice to enroll.[15] In sum, 4% of the class of 2011 dropped out without a degree, 17% completed the lower track, 36% the intermediary track, and 43% obtained a high school degree (Statistisches Bundesamt, 2013, p7). As a final remark, it is important to realize that in Germany, an apprenticeship degree has a high standing and a reputation similar to an university degree —especially when acquired after completing high school.

Summing up, in the subsequent analysis we distinguish between the four most commonly taken education paths in Germany, which we index by j. The student can choose to drop out ($j = 0$), invest in an apprenticeship directly after completing either the lower or intermediate school track ($j = 1$), or continue schooling in high school. After completing high school, the student can decide whether to invest in an apprenticeship ($j = 2$) or continue to university studies ($j = 3$).

Data sources

Our primary data source is the German Socio-Economic Panel [SOEP]. We focus on young adults, ages 16 to 17 years, who have newly entered the survey population by answering the youth questionnaire between 2000 and 2013. The SOEP is a household panel that provides a rich set of parental background information. We use all available waves of data collection to follow the young adults over time up to 14 years. Additionally, we combine the individual-level data with regional labor market information and educational supply and demand mea-

[14]Additionally, students could also drop out after completing high school, but this rarely occurs in practice (see also Fossen and Glocker, 2014). Note that here university subsumes universities of applied sciences. While it would be interesting to consider those separately, we have to leave this to future research due to our current sample size.

[15]We, therefore, include the indicator variable "In high school with 17" in all regressions. We also estimated the regressions of our main Table 1.3 separately —fully saturated in this variable— for the two groups and present the results in Appendix Table A.2.5.

sures based on 96 geographic regions, which we will refer to as Ror (for their German name *Raumordnungsregionen*).[16] All regional information is matched according to the individual's residency when answering the youth questionnaire, and lagged by one year to avoid endogeneity or reverse causality. Unless stated otherwise, we only use variables assessed in the youth questionnaire to avoid any biases from conditioning on outcomes (Angrist and Pischke, 2009, p64f).[17]

Sample selection

As stated above, we exclude all individuals who have already started an apprenticeship. Moreover, we exclude students with missing information in the core variables: subjective belief, GPA, and educational status. All other missing information are included along with corresponding indicator variables for missing observations. This selection results in a sample size of 3,610 individual observations. In the longitudinal analysis, we additionally require at least 2 years of information to assess the end of secondary education and the start of a post-secondary education (reducing the observations to 2,116), and to assess educational completion, we restrict the sample to students who responded for at least 5 years of data collection (1,372).[18]

Descriptive statistics

Our main variable of interest is the subjective completion belief, p_i, that was assessed by the following question:

> Think about your future in your job and private life: how probable is it, in your opinion, that the following events will occur?
>
> [Please check off a probability on the scale from 0% to 100%.]
>
> *You successfully finish your vocational training or university studies?*[19]

There are two caveats about how the question is assessed. First, the question is only elicited

[16] A map of the Ror's is provided in Appendix A.1, Figure A.1.1. The data source is INKAR 2012 provided by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR, 2013). For more information, see Pinger (2015) who also uses this additional data source. Moreover, we add the number of universities (higher learning institutions) as a proxy for distance to university provided by the statistical agency of Germany (Statistisches Bundesamt).

[17] We develop a structural model below in order to account for sequential decision-taking and to avoid associated biases.

[18] More information on missing values and the construction of the variables can be found in the Appendix, Table A.1.2.

[19] Students could answer on an eleven point scale. The exact wording in German is:
Wenn Sie sich einmal Ihre berufliche und private Zukunft vorstellen: Wie wahrscheinlich ist es, dass die folgenden Entwicklungen eintreten werden? [Stufen Sie bitte jeweils die Wahrscheinlichkeit auf einer Skala ein, die von 0 Prozent bis 100 Prozent geht.] *Ihre Ausbildung oder Ihr Studium erfolgreich abschließen?*

once. Second, the question does not elicit beliefs for every possible counterfactual education. We discuss the implications of these issues later in the estimation results. For now, we focus on the role of initial beliefs in the combined effect of any post-secondary education. In this way, the question directly relates to the outcomes that we assess. In Figure 1.2, we plot histograms of the subjective beliefs by students' intentions to invest in education. Intention to invest is a self-reported measure of educational aspiration that asks students to indicate which further educational degree (if any) they plan to complete. It is constructed analogously to our outcome variable: 0 refers to no further educational aspiration; 1, to apprenticeship; 2, to tertiary apprenticeship (high school and apprenticeship); and 3, to university studies.[20]

— — — Figure 1.2 about here — — —

Overall, the German students appear to be confident about finishing a post-secondary education, as most of the adolescents report a probability above 50%. The distributions of students implicitly aspiring to a high school degree (implied either by tertiary apprenticeship or university, Panels C and D) are very similar in shape. Yet, the subjective beliefs of young adults who aspire to a university degree are more concentrated and slightly shifted to the right compared to their high school counterparts who aim for a tertiary apprenticeship position. The mode of the distribution of students who intend to start an apprenticeship without finishing high school lies at 100%. Finally, students with no educational aspiration display a much larger spread in their beliefs. In what follows, it is important to keep in mind that most variation —and the bulk of the students' beliefs— are located between 60 and 100%.

Analogously, Table 1.1 presents descriptive statistics by aspiration level for our baseline sample.[21]

— — — Table 1.1 about here — — —

At the bottom of the table, we present the sample shares of the intentions to invest: Most students want to complete an apprenticeship, followed by university studies. A substantial share

[20] That means that, for consistency with our outcome variables, students who want to enroll in an apprenticeship first and then continue with supplementary courses that prepare for university are subsumed into the apprenticeship category. Moreover, we also cannot distinguish in detail between students who first want to complete an apprenticeship and then a high school degree, without aiming to go university. However, this path is neither optimal from a human capital investment perspective nor one that is commonly taken in Germany.

[21] More information on the construction of the variables can be found in the table notes. Unconditional descriptive statistics for the various subsamples considered in the analysis below are presented in the Appendix, Table A.1.1.

wants to complete a high school education and an apprenticeship (tertiary apprenticeship), and roughly 10% do not aspire to any professional education. It is reassuring that the sample statistics are broadly consistent with the population statistics presented above (Statistisches Bundesamt, 2013).

The individuals who aspire to a university education are on average the most confident about successfully completing their post-secondary education, and have the lowest standard deviation. However, all young adults who have any educational aspirations exhibit a similar level of completion beliefs —which are close to 80%— as opposed to those without educational aspirations. The fact that all students have positive beliefs, even the ones who do not plan to invest in further training or education, can be rationalized in a simple expected utility framework where students weight their utility from education by their beliefs about their completion probabilities and report their aspirations based on their highest expected utility. This is also the interpretation we pursue in the following analysis.

Some interesting patterns emerge when relating educational aspirations to our three measures of academic ability: Aspirations are increasing in the grade point average [GPA].[22] Prior track recommendations at the age of 10 years seem to be a good indicator for the aspirations up to 7 years later, which could either be caused by a well-working ability streaming or a manifestation of students' expectations as a result of early-tracking. Interestingly, having no educational aspirations occurs in all tracks, and the largest share of students without aspirations is found in high school. This could be explained by a default effect, as the survey elicits these aspirations at a time when students not enrolled in high school have to make an active decision as opposed to their high school counterparts who can follow their track and decide after obtaining a high school degree.

We assess the adolescents' personality by locus of control, risk attitudes, and the Big Five personality inventory.[23] Educational aspirations are positively associated with the locus of

[22]The GPA refers to the student's average of the German and Math grade, which is standardized over the sample population we present in our main results from Table 1.3. We also standardize the GPA within school track in Appendix Table A.2.3 to show that the choice of standardization does not drive the results. We further standardize GPA within federal states to show that different grading levels do not affect the results, see Appendix Table A.2.4.

[23]We standardize all the principal components of the personality variables (all but risk attitudes, which are assessed by one question only), small deviations from (0,1) result from the missing values which do not enter the standardization but are set to 0 afterwards. The locus of control has been developed by Rotter (1966), the Big Five inventory by Costa and McCrae (1992) and validated in the SOEP version by Hahn, Gottschling and Spinath (2012). Risk attitudes have been introduced and extensively studied by Dohmen et al. (2011) and references therein.

control, which measures to what extent a person believes her life is under her own control. Among the standard Big Five inventory, aspirations increase with openness, agreeableness, and extraversion but they are less monotonically related to conscientiousness or neuroticism. Unconditionally, aspiring to a university degree is positively associated with risk attitudes.

Individuals' characteristics and family backgrounds are captured by their gender, number of siblings, whether they are second-generation immigrants (persons whose parents were both born in a foreign country), whether at least one parent has a college education, is currently unemployed, and the logarithm of the net household income. Aspirations tend to be higher among males, children from smaller families, natives, persons with employed and college-educated parents or with a higher household income.

The regional labor and education market (Ror) characteristics relevant for the students' choices set are a mix of (exogenous) educational supply and demand shifters. We use the cyclical component of the youth unemployment rate, and the number of apprenticeship positions, students, high school graduates, and universities in the region. Throughout, the aspirations are increasing with the local labor and education market characteristics as expected; only children with no educational aspirations tend to have no clear ordering. In the following analysis, we will also account for region and year of first questioning (which is roughly identical to students' age).[24]

1.3 Determinants of subjective completion beliefs

To analyze how the variables we discussed in the previous section relate to subjective completion beliefs, we estimate OLS regressions of the model

$$p_i \quad = \quad x_i'\beta^p + v_i, \tag{1.1}$$

[24]For some of the regressions, the number of students within a state is too small. To obtain consistent samples, we use a broader grouping by dividing Germany into the following 5 regions (and an indicator for missing values). Southern Germany: Baden-Wuerttemberg, Bavaria; Eastern Germany: Berlin, Brandenburg, Saxony, Saxony-Anhalt, Mecklenburg-Western Pomerania; Central Germany: Hesse, Thuringia; Western Germany: North Rhine-Westphalia, Rhineland-Palatinate, Saarland; Northern Germany: Bremen, Hamburg, Lower Saxony, Schleswig-Holstein. We present analogous results of our main Table 1.3 in Appendix Table A.2.3 were we use federal states fixed effects, as the jurisdiction over educational policies are on the federal state level. The results are qualitatively the same.

where i indexes individuals, p_i is the subjective completion belief, x_i are varying sets of explanatory variables with corresponding vector of coefficients β^p, and υ_i is an unobserved error term.

The estimates are presented in Table 1.2.[25] In Column (1), the beliefs are explained solely by academic ability. In Column (2), we add the personality measures; in Column (3), individual and family and individual background characteristics; and, finally, in Column (4), regional measures, year and region fixed effects.

— — — Table 1.2 about here — — —

The explained variation, as measured by the adjusted R^2, increases substantially only when academic ability and personality measures are included, but stays relatively unaffected when adding individual and family characteristics, or fixed effects and regional characteristics. The joint significance tests for subsets of variables reported at the bottom panel of the table give analogous results: Academic ability and personality characteristics are highly significant across all regressions; individual and family characteristics are jointly significant; labor market characteristics, regional and time effects are not. Since the labor market coefficient estimates are neither jointly nor individually significant, we omitted these estimates from the table.

Looking at the determinants individually, all academic ability variables are consistently positive and significant. Somewhat surprisingly, already being enrolled in high school does not alter students' subjective completion beliefs. This might be due to this effect being conditional on track recommendation. As hypothesized by Coleman and DeLeire (2003), the locus of control is a very important determinant of subjective completion beliefs throughout the regressions, both in magnitude and significance.[26] Risk attitudes do not matter once family characteristics are accounted for. Our regressions indicate that among the Big Five measures of personality, conscientiousness is the most influential in shaping subjective beliefs. This finding highlights the importance of conscientiousness for educational outcomes, as is consistently found in the literature (see, *inter alia*, Borghans, Meijers and Ter Weel, 2006). While we find little evidence that openness or neuroticism influence completion beliefs, extraversion has a coefficient which

[25] Note that our dependent variable is a fraction. In the Appendix, Table A.2.1, we present fractional response regressions (as in Papke and Wooldridge, 1996, 2008). The results are virtually indistinguishable from the OLS estimates.

[26] Caliendo, Cobb-Clark and Uhlendorff (2015) also find a strong link between subjective beliefs and the locus of control in the realm of job search among the unemployed.

is about half as large as conscientiousness, and the effect of agreeableness is about half as large as extraversion.

On average, females seem to have lower subjective completion beliefs. This estimate is, however, only marginally significant (at least conditional on personality and academic ability). Household income is positively and significantly related to subjective completion beliefs. Being a second-generation immigrant is significantly negatively associated with subjective beliefs. However, the significance vanishes after including regional determinants. This suggests a segregation effect, with immigrants being located in less economically and educationally active areas. The other covariates are insignificant and mostly very small in magnitude.

1.4 Subjective completion beliefs and educational outcomes

In this section, we turn to our central question of how subjective completion beliefs measured at age 17 years relate to intended investments in education, actual investments in education, and, finally, educational degree attainment. To fix ideas, let the individual i's utility u_{ij} from choosing an uncertain post-secondary educational track ($j \geq 1$) be

$$u_{ij} = \begin{cases} \mu_{ij} + \varepsilon_{ij} & \text{with probability } p_{ij} \\ \bar{\mu}_{ij} + \varepsilon_{ij} & \text{with probability } (1 - p_{ij}) \end{cases}, \tag{1.2}$$

where p_{ij} is the subjective completion belief, μ_{ij} ($\bar{\mu}_{ij}$) is the utility from (not) completing, and ε_{ij} is an utility component unaffected by completion. The associated expected utility is

$$\begin{aligned} U_{ij} &= p_{ij}\mu_{ij} + (1 - p_{ij})\bar{\mu}_{ij} + \varepsilon_{ij} \\ &= \bar{\mu}_{ij} + p_{ij}(\mu_{ij} - \bar{\mu}_{ij}) + \varepsilon_{ij}. \end{aligned} \tag{1.3}$$

Hence, adolescents get a baseline utility from attending a particular educational track $\bar{\mu}_{ij}$. The subjective completion belief p_{ij} weights the utility differential between completing and not completing an educational track either up or down. Since not investing in an educational track does not involve educational uncertainty, its utility is simply

$$U_{i0} = \mu_{i0} + \varepsilon_{i0}, \text{ with certainty.} \tag{1.4}$$

In this section, we assess the investment in any post-secondary education $U_{ij} = U_i$ for $j \geq 1$, against not investing U_{i0}. The subjective belief p_i therefore corresponds directly to the question in the survey. A student prefers to invest in education if $U_i > U_{i0}$; where, by standard normalization, $\mu_{i0} = 0$. Taking averages across individuals, adding covariates x_i that measure observed preferences and skills, and assuming that $\nu_i = \varepsilon_i - \varepsilon_{i0}$ follows a standard normal distribution, we estimate probit models of the form

$$d_i = 1[\alpha p_i + x_i'\beta^d + \nu_i > 0]. \tag{1.5}$$

We consider three binary outcomes d_i. First, whether a student intends to invest in any further education, which is measured concurrently with subjective beliefs at age 17 years. Second, whether a student actually invests in any further education; that is, whether the student started an apprenticeship or, tertiary apprenticeship, or enrolled in a university. This event can be a few months or a few years away. Third, whether a student completes an apprenticeship or university degree, an event that is at least a couple of years away.

When d_i stands for the intention to invest, the expectation of (1.5) gives $P(U_i > 0)$, so that $\alpha = \mu - \bar{\mu}$. A similar interpretation is possible when d_i represents the second outcome. It then corresponds to the revealed preferences of actual investment in post-secondary education. The interpretation is somewhat different when d_i stands for the third outcome, the completion of a degree. In this case, α gives an indication of the student's ability to incorporate information beyond that in x_i into their forecast of $d_i = 1$. Here, we interpret the subjective beliefs in a similar vein as Finkelstein and McGarry (2006). The adolescents process all their available information in forming their beliefs, meaning that relevant information over and above their subjective beliefs are either not used, not used efficiently, or influence the decision through another channel than subjective completion uncertainty.

— — — Table 1.3 about here — — —

Table 1.3 contains the estimation results. Vertical panels (A) to (E) present the probit regressions of the subjective completion beliefs on the different educational outcomes. In each panel, we report the estimated coefficients, robust standard errors (in parentheses), average marginal effects (in squared brackets), pseudo R^2_n for the model estimated with and without p_i, and sample statistics for the respective subsamples. Columns (1) to (4) contain the simple probit estimates of the educational outcomes on the subjective beliefs and varying sets of covariates: The specification in Column (1) contains, apart from p_i, only an indicator of whether the student is currently in high school, region and year fixed effects. Thus, in this specification, any other variable acts on the intention to invest in education through its effect on p_i. The next columns progressively control for the sets of academic (Column 2), personality (Column 3), and family and labor market variables (Column 4). We turn to the results in Columns (5)-(8) at the end of this section.

Panel (A) contains results corresponding to the intention to invest in any post-secondary education. Unsurprisingly, uncertainty appears to be important for aspirations: The coefficients on subjective beliefs are large and highly significant throughout the probit regressions. The average marginal effects are economically relevant. In the most parsimonious specification, increasing the subjective beliefs in the population by one standard deviation increases intentions to invest in post-secondary education by 2.7 percentage points (0.14×0.198), which is very large given that only 9.2% of students do not intend to invest in a post-secondary education. This changes little if we condition on increasing sets of background characteristics commonly considered in the literature. A one standard deviation change results in an increase of 2.1 percentage points using all background characteristics. Moreover, the increase in the pseudo R^2 when including the subjective beliefs is similar to the increase when adding both personality and family background. Thus, we find subjective beliefs are strongly related to intended behavior, a result consistent with Huntington-Klein (2015*b*).

Several explanations can account for this observed correlation between beliefs and aspirations; therefore, we examine whether the link from beliefs to intention carries over to revealed preferences in actual investments (at least two years later). Our dependent variable is now an indicator that equals one if the student started any post-secondary education. Panel (B) uses the broadest sample possible for this question. Compared to (A), it only excludes students who are still in school and students who have not completed any subsequent questionnaires two years after the baseline questionnaire at age 17. The average marginal effect is somewhat smaller than for the intentions, ranging from 1.4 to 0.9 percentage points for one standard deviation increase in the subjective beliefs.[27] In Panel (C), we show that similar results are obtained when restricting the sample to students who expressed earlier intentions to invest in post-secondary education. This shows that the subjective completion belief drives not only hypothetical, intended investment, but it also has real behavioral consequences. Compared to the previous results on students' intentions-to-invest, the set of family background and labor market variables explain a larger fraction of the completion belief effect and exhibit a substantial explanatory power.

It is interesting to compare how the subjective completion beliefs relate to actual completion (at least five years later). This can be interpreted as how well the students can predict their future outcomes.[28] The estimation results are given in Panels (D) and (E). The average completion rate is roughly 55%. Unconditionally, a one standard deviation increase in the subjective beliefs increases completion rates in the population by 3.3 percentage points in the overall sample (Panel D), and by 3.6 when conditioning on the sample with positive intentions (Panel E). This decreases to 2.6 and 3.3 percentage points, respectively, when including the full set of individual, family, and regional characteristics. Comparing the coefficients across rows, a notable result is that for degree completion, the set of covariates that affect the coefficient of p_i most is that of the personality measures. Including these variables reduces the estimated coefficient by about 15 to 20 percent. The explanatory power of personality, family background, and labor market characteristics are substantial. This suggests that the students do not optimally account for this information when forming their beliefs. Again, the explanatory power of the beliefs is substantial.

Taken together, the results show that subjective completion beliefs formed during secondary education are predictive over a long time horizon for future post-secondary education. The

[27] Almost all adolescents in Germany start some post-secondary education, 95.6% in our sample, which explains why the average marginal effect for investment is smaller than that for intentions despite an estimated coefficient of similar magnitude.

[28] Since completing a program and graduating takes some time, we only consider students which we see at least five years after they have taken the youth questionnaire when they indicated their completion beliefs. This further reduces our available sample. Moreover, it is clear that students who were interviewed in earlier years are more likely to have completed their degrees simply by virtue of being in the sample for a longer period of time. However, this mechanism is captured by the year fixed effects, and is therefore unlikely to bias our results. A second concern is that some of the observations are censored: As of the time we observe them, some students have not yet completed their degree, but they might do so in the future. In this sense, our results should be interpreted as representing the average effect of completion beliefs on completion within a given time frame.

subjective beliefs are predictive even after accounting for a large set of previously identified, important characteristics. In the appendix, we present further results showing the robustness of these findings across a number of alternative specifications. We show that the results are robust to dichotomizing the subjective beliefs to a dummy variable, thus accounting for potential non-linearity as discussed in Pinger (2015). Further, since academic ability is found to be the main determinant of learning about one's own ability in the literature, we use various reasonable standardizations of GPA that account for potential differences in grading across federal states, or within high school versus no high school. We also use a fifth-order polynomial to show that the beliefs do not pick up non-linearities in academic ability. Additionally, we use federal state dummies instead of the region dummies used in the main specification (cf. footnote 24). Finally, we present separate estimations for students enrolled in high school when answering the youth questionnaire to account for the different default choices discussed above in a completely flexible way.

A remaining concern might be that the uncertainty is confounded with unobserved heterogeneity. We therefore use a bounding strategy for the coefficients by taking potential selection on unobservable tastes and preferences for education into account when estimating (1.5). We use the approach developed by Altonji, Elder and Taber (2005*a*,*b*, 2008, hereafter, AET). More specifically, we simultaneously estimate the models given in (1.1) and (1.5), imposing the following dependence between the error terms:

$$(\nu_i, \upsilon_i) \sim \Phi_2(0, 0, 1, 1, \rho), \tag{1.6}$$

where $\Phi_2(\cdot)$ denotes the bivariate normal distribution, and its arguments are the two errors' means, variances, and their correlation. In other words, we estimate probit models for all outcomes d_i with p_i as a normal endogenous explanatory variable [denoted probit eev hereafter].[29,30] The bounding is achieved by setting the correlation coefficient ρ to increasing values

[29] The corresponding log-likelihood is given by

$$\ln L(d_i, p_i; x_i, \alpha, \beta^d, \beta^p, \rho) = \sum_{i=1}^{n} \ln \Phi \left[(2d_i - 1) \left(\frac{x_i'\beta^d + \alpha p_i + \rho(p_i - x_i'\beta^p)}{\sqrt{1-\rho^2}} \right) \right] + \ln \phi\left(p_i - x_i'\beta^p\right). \tag{1.7}$$

For more information, see the discussion in Greene (2012, p747f).

[30] In contrast to AET, our main variable is a fraction rather than an indicator. Instead of estimating a bivariate probit, we therefore estimate a probit eev. The use of a continuous normal variable is motivated by the estimation of (1.1), where we found that it made little difference whether it was estimated by OLS or a

until the coefficient of the subjective beliefs α tends to zero. Note that Column (4) in combination with Table 1.2's Column (4) is equivalent to the probit eev with $\rho = 0$. AET argue that the selection-on-observables is a reasonable (upper) bound on the selection-on-unobservables. Therefore, we also estimate the model replacing

$$\rho = \frac{cov(x_i'\beta^d, x_i'\beta^p)}{var(x_i'\beta^d)} \equiv \hat{\rho}^o$$

as a suggestive upper bound. Columns (5) to (7) contain the probit eev estimates using the full set of covariates and ρ constrained to 0.1, 0.3, and 0.5. Finally, Column (8) constrains ρ to be equal to the selection-on-observables $\hat{\rho}^o$. Up to a correlation of 0.3 all coefficients are positive, and for aspirations and intentions they remain statistically significant. This is a sizeable correlation when comparing it to the applications considered in AET. When using the AET bound of selection-on-observables in the last column, the coefficients are all statistically significant and similar in magnitude to those using all covariates and a correlation between 0 and 0.1. This indicates that the results are robust to a sizeable selection-on-unobservables.

The results presented in this section indicate that the uncertainty of 17-year-old students about completing an educational degree is an important determinant of educational choices and outcomes. While we focused our discussion on the average effect, another effect which is of interest is the one corresponding to the marginal student (a student with an outcome probability of 50%). An increase in the subjective belief by one standard deviation (0.2) for this student would: increase her intention to invest by 5.7 percentage points (i.e., $\Phi(0.2 \times 0.716) - \Phi(0))$, her investment by 6.9 percentage points (5.8 if she stated an intention to invest), and her completion by 2.6 percentage points (3.4 if she stated an intention to invest). In sum, differences in beliefs about being capable of successfully finishing a post-secondary educational degree can explain not only differences in intended future investments in schooling, but in actual investments as represented by enrollment into university or obtaining an apprenticeship position. Moreover, students with higher subjective beliefs are also associated with higher completion rates, even after controlling for several potential confounders and allowing for some

fractional response model (cf. Appendix Table A.2.1). As a robustness check we dichotomize the subjective beliefs at $p \geq 70\%$ and estimate bivariate probit regressions as in AET. Estimates for such an approach can be found in the Appendix Table A.2.2. The results are similar but somewhat more conservative, possibly due to the reduced variation.

selection-on-unobservables.

1.5 A view at the disaggregated level

To understand how subjective beliefs influence educational choices and outcomes, we proceed with a more disaggregated analysis: different educational choices. In the estimations before, we implicitly assumed the subjective belief measure has the same effect on all educational investments. Clearly, while this is a useful simplification that allows to gauge overall average effects, it might also hide important differences in how completion beliefs explain, say, enrollment in a university program versus enrollment in a vocational training degree. In this section, we separately asses each of the three educational investments: apprenticeship ($j = 1$), tertiary apprenticeship ($j = 2$), and university ($j = 3$). Compared to the previous sections, in which the subjective belief measure corresponded directly to the outcome, for the disaggregated educational tracks one would ideally like to assess the role of counterfactual choices. Unfortunately, these were not elicited in the survey. Therefore, we condition on students' aspirations; however, results have to be interpreted more cautiously.

As before, the analysis starts at the level of intended investment. This time we present results from a multinomial probit model with four outcomes. The base category is not having any intention to invest, and the remaining categories are the intention to invest in each of the three educational choices mentioned above. The results are presented in Table 1.4, whose four columns represent specifications with the increasing sets of covariates discussed previously. We present the χ^2-statistic and corresponding p-value for the likelihood ratio test against a restricted model without subjective beliefs.

— — — Table 1.4 about here — — —

All coefficients are statistically significant and large in magnitude. The first column shows that the overall average marginal effect of subjective beliefs (in squared brackets) found before stems almost exclusively from the fact that students with higher completion beliefs aspire to a university education. Notice that the sum of the three average marginal effects is roughly equal to the corresponding effect presented in the previous section.[31]

[31] We omit the average marginal effect for the base category: it is equal to minus the sum of the marginal effects of the remaining categories.

Yet, the second column shows that academic variables, such as GPA, are central in shaping this effect. Once the academic background has been accounted for, the effect of subjective belief works mainly through the apprenticeship channel. The average marginal effect for university is reduced substantially and rendered statistically insignificant. Thus, high GPA has the effect of inducing high completion beliefs, which in turn pushes students towards desiring a university degree. But within a given GPA level, a higher completion belief is positively associated with starting an apprenticeship. Compared to this big shift, the changes resulting from adding personality, family background, and labor market variables are modest (at least beyond its effect through GPA or personality skills). This result is in line with those found in the literature that for college students most of the information is based on measures of academic ability (Zafar, 2011*b*; Stinebrickner and Stinebrickner, 2012, 2014*b*; Milla, 2014). Yet, for those who choose a less theoretical education, other characteristics seem to be more influential.

Turning to the behavioral responses, we estimate separate regressions for the three subsamples according to intended educational choice. That is, we address the question, for example, of how does a higher subjective completion belief increase a student's university enrollment and completion probabilities, given that the student aspired to a university degree. Estimates are reported in Table 1.5. Columns (1) to (4) display the results for enrollment and (5) to (8) for completion. The three vertical Panels (A) to (C) contain separate probit regressions of each educational track.

— — — Table 1.5 about here — — —

Panel (A) reports the estimates for investment and completion of an apprenticeship. The results indicate that the effects of subjective beliefs are indeed large and statistically significant, with the average marginal effect ranging from about 10 to 15 percentage points, depending on the set of control variables used. With 883 observations, the subsample with intended investment in apprenticeship is the largest of the three subsamples, accounting for almost half of the total number of observations. The subsample for intended tertiary apprenticeship has only 456 observations. In three out of the four specifications shown in the middle panel, the estimated coefficient is insignificant. The coefficient —and hence the marginal effect— increases as more covariates are controlled for, and only reaches marginal significance in the last column of the panel. The bottom panel, containing the results of the subsample aspiring to a university education, features the opposite pattern. Here, completion beliefs have a large, statistically significant effect on enrollment in university. However, academic background explains almost half of the effect. Adding more sets of control variables further erodes the effect of subjective beliefs on college enrollment.

Comparing the results across the three panels, it appears that the decision to enroll in a post-secondary educational program is related most strongly to subjective completion beliefs for those students aspiring to a university degree (see columns 1 across panels). At the same time, the determinants of these beliefs are mainly related to observable academic and demographic variables for the university-aspiring students. For the two apprenticeship streams, the observable academic and demographics add comparatively little information to the completion beliefs. Moreover, for the tertiary apprenticeship, there even seems to be a negative correlation with these characteristics, but the estimation results are too imprecise to allow for further interpretation.

To conclude this part, we estimate analogous probit regressions for the probability to graduate. The results are depicted in the right-hand-side panel of Table 1.5. Small sample size issues are a concern, especially for the tertiary apprenticeship graduation regressions. However, the estimates are consistent with our previous results. In particular, the aggregate effect found in the previous section is corroborated in the apprenticeship category. For students who indicated their intentions to invest in an apprenticeship degree at age 17 years, the subjective completion beliefs are highly informative about their actual completion years later. The average marginal effect is close to 30 percentage points —a figure that is reduced to about 20 percentage points after accounting for differences in observables. We cannot estimate precise effects for tertiary apprenticeships. While the point estimates are sizeable, none of them are statistically significant. The results for students aspiring to a university education also echo the previous results. The effect of the completion beliefs at age 17 years is large and statistically significant: a one-standard-deviation change in p_i increases the probability of graduating from university by about 6 percentage points for a student with a baseline graduation probability of 50 percent. Finally (and as before), the available control variables, particularly academic background variables, explain a large portion of this effect.

1.6 A dynamic model of educational choice

In this section, we conclude our investigation by developing and estimating a model of educational investment along the lines of Taber's (2001) seminal contribution that encompasses three features. First, we allow for the sequential nature of the process: students can only decide whether they want to go to university if they chose to finish high school previously (Comay, Melnik and Pollatschek, 1973; Altonji, 1993). Second, we introduce the dynamics of the optimization process: when deciding whether to go to the labor market or to go to high school, forward-looking students account for the option value of continuing education after finishing high school (Stange, 2012; Trachter, 2015). Finally, we allow for unobserved factors that influence student utilities derived from their choices, which may be correlated across choices and over time, a topic of substantial attention in the returns to education literature (see, e.g., Card, 2001; Belzil, 2007, and references therein).

Model

We consider a stylized two-period model in which students sequentially choose between risky educational paths, as outlined in Figure 1.3.

— — — Figure 1.3 about here — — —

Ex ante, students do not know for certain whether they will successfully complete the chosen education track, but they have subjective beliefs, p_i, about finishing. The first period or first stage ($T = 1$) occurs when students finish compulsory education at the age of 17 years. At this point, they face the choice between dropping out of school ($d_{i1} = 0$), investing in an apprenticeship training ($d_{i1} = 1$), or continuing with high school education ($d_{i1} = 2$). A high school degree involves the option value of continuing with tertiary education. Students who choose high school reach the second period ($T = 2$), where they graduate from high school and now have the choice of either investing in a tertiary apprenticeship ($d_{i1} = 2, d_{i2} = 0$) or in a university education ($d_{i1} = 2, d_{i2} = 1$).[32,33]

[32] As noted before, students could also drop out at this point, but this is an extremely rare event in the data and therefore not modeled (see also Fossen and Glocker, 2014).

[33] We only focus on the initial beliefs in shaping young adults' educational choices because we lack a repeated measurement of the subjective beliefs at the end of high school that would allow us to study the learning about ones' own ability in more detail.

As mentioned in Section 1.4, apprenticeships, tertiary apprenticeships, and university all involve uncertainty, which we model according to equations (1.2)-(1.4). A key assumption of this approach is that utility can be decomposed into a component that depends on the realization of graduation (μ_{ij} for graduation vs. $\bar{\mu}_{ij}$ else) and an idiosyncratic component unaffected by graduation, ε_{ij}, which captures features such as a preference for attending, say, university irrespective of receiving a degree.

By backward induction, we begin with the students' second stage problem. Students advancing to the second stage choose between starting a tertiary apprenticeship ($j = 2$) or going to university ($j = 3$). We denote this choice by d_{i2}, a binary variable where 1 represents choosing university,

$$d_{i2} = \begin{cases} 1 & \text{if } U_{i3} - U_{i2} > 0 \\ 0 & \text{if } U_{i3} - U_{i2} \leq 0 \end{cases}$$

which we specify analogously to equation (1.4) by

$$U_{i3} - U_{i2} \;\; = \;\; \alpha_3 p_i + x'_{i,t+1}\beta_3 + \delta_3\theta_i + \nu_{i3} \equiv z_{i3,t+1} + \nu_{i3}.$$

Here, $\alpha_3 = (\mu_3 - \mu_2) - (\bar{\mu}_3 - \bar{\mu}_2)$, $\nu_{i3} = \varepsilon_{i3} - \varepsilon_{i2}$, and $x_{i,t+1}$ consists of the same covariates considered above, although we include time-varying labor market conditions measured two years after answering the youth questionnaire, which is the time one would need to start a higher education after obtaining a high school degree.[34] This exogenous variation in the decision problem induced by the timing of the events provides an additional source of identification (see, Taber, 2000; French and Taber, 2011, for a discussion on the identification for these models), which has become standard practice in the literature on dynamic models of educational choice (e.g., Taber, 2001; Heckman et al., 2014). To allow for dependence of the unobservables between the two time periods in a flexible way, we add a standard normal random variable θ_i to the utilities, capturing unobserved tastes and preferences for education. We assume that $\nu_{i3} \sim N(0, \sigma_3)$, thus the probability of choosing university relative to tertiary apprenticeship is

[34] We use students' location at the age of 17 for the region regional variables, to avoid a bias due to moving.

given by

$$P(d_{i2}=1) \quad = \quad \Phi\left(\frac{z_{i3,t+1}}{\sigma_3}\right),$$

where $\Phi(\cdot)$ represents the univariate normal cdf.

In the first stage, the student has an expectation about her second stage decision (she knows the distribution of ν_{i3}) but does not know her realized value. If students knew their realized ν_{i3} at the time of the first stage, the model would reduce to a simple static polychotomous choice problem, similar to those estimated and reported in Table 1.4. Thus, the students' expectation about her value of advancing to the second stage, as formed during the first stage, is

$$\mathrm{E}(\max(U_{i3}-U_{i2},0)) \quad = \quad \sigma_3\left[\Phi\left(\frac{z_{i3,t}}{\sigma_3}\right)\frac{z_{i3,t}}{\sigma_3}+\phi\left(\frac{z_{i3,t}}{\sigma_3}\right)\right] \equiv EV_i,$$

and $\phi(\cdot)$ denotes the normal pdf. Now the labor market and educational supply and demand characteristics are measured at time t, one year before the adolescent answers the youth questionnaire corresponding to her information set. The difference between high school and drop out utility is then

$$\begin{aligned} U_{iHS}-U_{i0} &= \alpha_2 p_i + x_{i,t}'\beta_2 + \delta_2\theta_i + EV_i + \nu_{iHS} \\ &\equiv z_{iHS,t} + \nu_{iHS}, \end{aligned}$$

which comprises EV_i, the option value of continuing to the second stage. In these types of models, we cannot distinguish between the baseline utility of the second stage tertiary apprenticeship and the utility of high school. The coefficients $\alpha_2, \beta_2, \delta_2$ thus capture the sum of these two effects (while the coefficients in $z_{i3,t}$ correspond to the differences between preferences for university and tertiary apprenticeship).

The apprenticeship utility is

$$U_{i1}-U_{i0} \quad = \quad \alpha_1 p_i + x_{i,t}'\beta_1 + \theta_i + \nu_{i1} \equiv z_{i1,t} + \nu_{i1},$$

where we set $\delta_1 = 1$, a necessary normalization to identify the impact of unobserved heterogeneity, θ_i, on latent utilities. By the bivariate normal assumption on the ν's we can write the probabilities

$$\begin{aligned} P(d_{i1}=2) &= \Phi_2\left(z_{iHS,t},\;\; z_{iHS,t}-z_{i1,t},\;\; 0.5\right), \\ P(d_{i1}=1) &= \Phi_2\left(z_{i1,t},\;\; z_{i1,t}-z_{iHS,t},\;\; 0.5\right), \\ P(d_{i1}=0) &= 1-P(d_{i1}=1)-P(d_{i1}=2). \end{aligned}$$

The individual likelihood contribution, conditional on the unobserved heterogeneity θ_i, is given by

$$\begin{aligned} l_i(\theta_i) = \; & \{1-P(d_{i1}=1)-P(d_{i1}=2)\}^{\mathbf{1}(d_{i1}=0)} \quad \times \quad \{P(d_{i1}=1)\}^{\mathbf{1}(d_{i1}=1)} \\ & \times \quad \{P(d_{i1}=2)[1-P(d_{i2}=1)]\}^{\mathbf{1}(d_{i1}=2,d_{i2}=0)} \\ & \times \quad \{P(d_{i1}=2)P(d_{i2}=1)\}^{\mathbf{1}(d_{i1}=2,d_{i2}=1)}, \end{aligned} \tag{1.8}$$

and to obtain the marginal likelihood contribution, we integrate over the distribution of θ,

$$l_i = \int l_i(\theta_i)\phi(\theta_i)d\theta_i,$$

an expression which we approximate by simulation, $\tilde{l}_i$, by taking random draws from the distribution of θ_i. We then maximize the simulated sample log-likelihood $\sum_i \ln \tilde{l}_i$.

Results

The estimation results are depicted in Table 1.6 in two panels. The left-hand-side panel contains estimates from a constrained version of the model without heterogeneity ($\theta_i = 0$ for all i), whereas the right-hand-side panel contains estimates from the full model with unobserved heterogeneity. Moving from left to right, the columns again contain the expanding set of covariates considered previously. With the exception of the local labor and education market characteristics, all the regressors are time-invariant.[35]

— — — Table 1.6 about here — — —

Table 1.6 depicts large and significant estimates for the coefficients of the subjective probabilities in the indices for both $d_{i1}=1$ and $d_{i1}=2$. Thus, these results, too, suggest that a higher

[35] The estimate of σ_3 is only identified when time-varying covariates are included (Taber, 2000; French and Taber, 2011). We present it therefore only in Columns (4) and (8).

p_i pushes students away from leaving school without further investments. In particular, the coefficients for $d_{i1} = 2$ suggest that subjective completion beliefs are important determinants of second-stage participation; of completing high school and beginning a tertiary apprenticeship or university studies (which confirms and extends the results found by Pinger, 2015). On the other hand, the coefficients for university are insignificant throughout, and close to zero when accounting for covariates. This indicates that, once in the second stage, the initial subjective beliefs are not informative about the choice of tertiary apprenticeship versus university. A potential explanation for this is belief updating in response to new information revealed by high school grades.

— — — Figure 1.4 about here — — —

Figure 1.4 illustrates the role of the option value, EV_i, in shaping the choice probabilities. The figure uses predicted probabilities obtained from the estimated parameters in Column (5) and evaluated at sample means. The left-hand-side panel artificially sets the expected value to zero; that is, we evaluate a constrained model where students ignore the option value of further investment. Thus, we interpret the coefficients from $d_{i1} = 2$ as corresponding only to high school utility, and we assume students neglect the option value of continuing to the second-stage choices. As expected, it can be seen by contrasting the two panels in the figure that the option value decreases the level of the apprenticeship probabilities and increases those of the second-stage choices. But the option value also affects how the probabilities change with the subjective belief, making the gradient on apprenticeship flatter —and even slightly negative for high values of p_i— and the gradient on university steeper. Therefore, the option value in conjunction with the subjective beliefs can play a substantial role in shaping the adolescents' high school investment.[36]

We now turn to the role of the unobserved preferences for post-secondary education or unobserved skills, θ_i. Comparing the two panels of Table 1.6, we see that all the significant coefficients in the right-hand-side panel, which accounts for such heterogeneity, are somewhat larger than the ones from the left-hand-side. Recall that θ_i has no natural scale, its scale has been fixed such that a unit coefficient in the index corresponds to apprenticeship. The

[36] The extent of these effects depends on the values of the covariates, which in Figure 1.4 were set to sample means. In Figure A.1.2 in the Appendix we present graphs where we set all the linear indices $x_i'\beta_j = 0$ and thus obtain effects which are much stronger.

coefficient on the linear index for the baseline second-stage utility is about 0.85 across all specifications (5)-(8) and highly significant. It shows that there is a strong positive correlation between unobserved preferences for apprenticeship and for high school. Unobserved preferences for education are very important for the adolescents investment decisions, as found in the prior literature (e.g. Bulman, 2015; D'Haultfoeuille and Maurel, 2013; Huntington-Klein, 2015*a*; Wiswall and Zafar, 2015*a*). Yet, there is no evidence for differences between unobserved tertiary-apprenticeship-specific skills versus university-specific skills, with the estimated coefficient being virtually zero, potentially a result of preference updating within high school.

Figure 1.5 further uses the results from Table 1.6 to visualize how the effect of p_i might differ for different "types" of students.

— — — Figure 1.5 about here — — —

Based on the estimates of the full specification from Column (8), we define four types by their academic ability level (high versus low GPA) and their unobserved skill level (high versus low θ_i) and plot their predicted choice probabilities against p_i, evaluated at sample means.[37] For students who have high observed and unobserved skills, subjective completion beliefs have negligible effects on investment probabilities. Yet, for adolescents with low unobserved skills (and high GPA), subjective beliefs positively influence all educational tracks. For students with low academic performance, subjective beliefs are more relevant if they have a low preference for education. It is also interesting to note that high GPA (for given level of θ_i) has a much larger effect on investment than high unobserved skills and preferences (for a given level of GPA). This suggests that the subjective beliefs are most relevant for students with low unobserved skills.[38]

In sum, the results from the dynamic sequential model with unobserved heterogeneity shed light on some aspects of educational choice which were masked in the reduced form models of Sections 1.4 and 1.5. One such aspect is that the sequentiality of choices shows that p_i has a highly significant effect on the combined high school and second-stage choice; in contrast, it was difficult to estimate precise effects for p_i in the static reduced form model where all four choices were disaggregated. Furthermore, we have seen that accounting for the option value and for

[37] Specifically, we define high and low values of GPA and θ_i as $\Phi^{-1}(0.75)$ and $\Phi^{-1}(0.25)$.

[38] Figure A.1.3 contains a similar graph to Figure 1.5, but evaluated at $x_i'\beta_j = 0$, thus yielding even more pronounced effects.

unobserved skills can modify the effect of p_i on the choice probabilities. Speaking more broadly, the structural estimates confirm the main results from the reduced-form estimations presented previously: subjective probabilities contain predictive information for educational investments even after accounting for differences due to an extensive set of controls and unobserved heterogeneity. Additionally, and consistent with the recent literature, throughout the analysis GPA has been shown to be the main driver of subjective beliefs. Thus, learning about one's own ability is largely determined by school grades already before entering a post-secondary education.

1.7 Conclusion

In this paper, we investigated the role of uncertainty for 17-year-olds on their post-secondary educational outcomes by means of subjective beliefs. Two features of this problem are the young age of the students at the time their subjective completion probabilities were elicited and the long time horizon of the choices to which these measures referred. Both features make this a difficult problem, and it is remarkable that these necessarily crude initial beliefs retain their predictive power over a period of several years. The effects of subjective beliefs on investment intentions and actual investments in any post-secondary education are substantial, remaining so even after controlling for observables. Moreover, subjective beliefs have explanatory power comparable to that of academic and personality variables combined. For the marginal student, a one standard deviation increase in subjective beliefs is associated with a 6 percentage points increase for investment intentions and a 7 percentage points increase for actual enrollment. Finally, the subjective probabilities of completion are also predictive of actual completion, increasing completion probability by 3 percentage points.

When disaggregating the educational tracks and estimating a structural choice model, we find the subjective beliefs most relevant for students who aim for a university degree. This is due to the information revealed by GPA, which broadly confirms results found in the literature. Most notably, we confirm Zafar's (2011*b*) finding that *ex ante* subjective beliefs continue to be important even until the degree is completed. Advancing his findings, we conclude that this is even true for subjective beliefs formed already in or before investing (or staying) in high school. Conditional on the academic ability, the subjective beliefs are most relevant for students who start an apprenticeship, which is largely driven by students with additionally low unobserved skills or preferences for post-secondary education. The literature on subjective beliefs in educational choice has largely ignored these students and evidence on their learning/decision-making processes are almost non-existent. Our study suggests that these students deserve more attention.

References

Almlund, Mathilde, Angela Lee Duckworth, James Heckman and Tim Kautz. 2011. Personality Psychology and Economics. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 1–181.

Altonji, Joseph G. 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics* 11(1):48–83.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*a*. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources* 40(4):791–821.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*b*. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113(1):151–184.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2008. "Unobservables Assess Bias from Using Selection on Observed Variables to when Evaluating Swan-Ganz." *American Economic Review: Papers and Proceedings* 98(2):345–350.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press, New Jersey.

Arcidiacono, Peter, Joseph V. Hotz, Arnaud Maurel and Teresa Romano. 2014. "Recovering Ex Ante Returns and Preferences for Occupations using Subjective Expectations Data." *NBER Working Paper No. 20626*.

Arcidiacono, Peter, Joseph V. Hotz and Songman Kang. 2013. "Modeling college major choices using elicited measures of expectations and counterfactuals." *Journal of Econometrics* 166(1):3–16.

Attanasio, Orazio P. and Katja M. Kaufmann. 2014. "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender." *Journal of Development Economics* 109:203–216.

BBSR, Bundesinstitut für Bau-Stadt-und Raumforschung. 2013. "Raumordnungsregionen (ROR)." Accessed 18.12.2015.
www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Raumordnungsregionen/raumordnungsregionen_node.html

Belzil, Christian. 2007. "The return to schooling in structural dynamic models: a survey." *European Economic Review* 51(5):1059–1105.

Borghans, Lex, Angela L. Duckworth, James J. Heckman and Bas ter Weel. 2006. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43(4):972–1059.

Borghans, Lex, Huub Meijers and Bas Ter Weel. 2006. "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry* 46(1):2–12.

Bound, John and Sarah E. Turner. 2011. Dropouts and Diplomas. The Divergence in Collegiate Outcomes. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 573–613.

Bulman, Geogre. 2015. "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics* 7(4):1–36.

Caliendo, Marco, Deborah A. Cobb-Clark and Arne Uhlendorff. 2015. "Locus of Control and Job Search Strategies." *Review of Economics and Statistics* 97(1):88–103.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5):1127–1160.

Carneiro, Pedro, Karsten T. Hansen, James J. Heckman and Pedro Carneiro. 2003. "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice." *International Economic Review* 44(2):361–422.

Cebi, Merve. 2007. "Locus of Control and Human Capital Investment Revisited." *Journal of Human Resources* 42(4):919–932.

Christofides, Louis N., Michael Hoy, Joniada Milla and Thanasis Stengos. 2015. "Grades, aspirations and post-secondary education outcomes." *Canadian Journal of Higher Education* 45(1):48–82.

Coleman, Margo and Thomas DeLeire. 2003. "An Economic Model of Locus of Control and the Human Capital Investment Decision." *Journal of Human Resources* 38(3):701–721.

Comay, Yochanan, Arie Melnik and Mosche A. Pollatschek. 1973. "The Option Value of Education and the Optimal Path for Investment in Human Capital." *International Economic Review* 14(2):421–435.

Costa, Paul Jr. and Robert McCrae. 1992. "Four ways five factors are basic." *Personality and Individual Differences* 13(6):653–665.

Cunha, Flavio and James J. Heckman. 2007. "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling." *Labour Economics* 14(6):870–893.

Cunha, Flavio, James J. Heckman and Salvador Navarro. 2005. "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers* 57(2):191–261.

D'Haultfoeuille, Xavier and Arnaud Maurel. 2013. "Inference on an extended Roy model, with an application to schooling decisions in France." *Journal of Econometrics* 174(2):95–106.

Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100(3):1238–1260.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9(3):522–550.

Finkelstein, Amy and Kathleen McGarry. 2006. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review* 96(4):938–958.

Foley, Kelly, Giovanni Gallipoli and David A. Green. 2014. "Ability , Parental Valuation of Education and the High School Dropout Decision." *Journal of Human Resources* 49(4):906–944.

Fossen, Frank M. and Daniela Glocker. 2014. "Stated and revealed heterogeneous risk preferences in educational choice." *IZA DP No. 7950.*

French, Eric B. and Christopher R. Taber. 2011. Identification of models of the labor market. In *Handbook of Labor Economics*, ed. Ashenfelter Orley and David Card Vol. 4 Part A Elsevier B.V. pp. 537–617.

Fryer, Roland G. Jr. 2013. "Information and Student Achievement: Evidence from a Cellular Phone Experiment." *NBER Working Papers no. 19113.*

Greene, William H. 2012. *Econometric Analysis.* 7 ed. Pearson Education Limited. Edinburgh Gate, Harlow.

Hahn, Elisabeth, Juliana Gottschling and Frank M. Spinath. 2012. "Short measurements of personality - Validity and reliability of the GSOEP Big Five Inventory (BFI-S)." *Journal of Research in Personality* 46(3):355–359.

Hastings, Justine S., Christopher A. Neilson, Anely Ramirez, Seth D. Zimmerman and Seth D. Zimmermann. 2015. "(Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data." *NBER Working Papers no. 21330.*

Heckman, James J., John E. Humphries, Gregory Veramendi and Sergio Urzúa. 2014. "Education, Health and Wages." *NBER Working Papers no. 19971.*

Huntington-Klein, Nick. 2015*a*. "Educational Choice: Subjective Data and Consumption Value." *Mimeo.*

Huntington-Klein, Nick. 2015*b*. "Subjective and projected returns to education." *Journal of Economic Behavior & Organization* 117:10–25.

Huntington-Klein, Nick. 2015*c*. "(Un)informed College and Major Choice": Verification in an Alternate Setting." *Mimeo.*

Hussey, Andrew J. and Omari H. Swinton. 2011. "Estimating the ex ante Expected Returns to College." *American Economic Review: Papers and Proceedings* 101(3):598–602.

Jensen, R. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125(2):515–548.

Keller, Tamás and Guido Neidhöfer. 2014. "Who Dares, Wins? A sibling analysis of tertiary education transition in Germany." *SOEPpapers 713.*

Kerr, Sari P., Tuomas Pekkarinen, Matti Sarvimaki and Roope Uusitalo. 2015. "Post-Secondary Education and Information on Labor Market Prospects: A Randomized Field Experiment." *IZA DP No. 9372.*

Light, Audrey and Wayne Strayer. 2000. "Determinants of College Completion: School Quality or Student Ability?" *Journal of Human Resources* 35(2):289–308.

Manski, Charles F. 1989. "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon." *Economics of Education Review* 8(4):305–312.

Manski, Charles F. 1993. Adolescent econometricians: How do youth infer the returns to schooling? In *Studies of supply and demand in higher education*, ed. Charles T. Clotfelter and Michael Rothschild. Chicago: University of Chicago Press pp. 43–60.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72(5):1329–1376.

Mazza, Jacopo. 2014. "Does Risk Matter? A Semiparametric Model for Educational Choices in the Presence of Uncertainty." *Mimeo.*

Milla, Joniada. 2014. "Tertiary Education Choices under Uncertainty." *Mimeo.*

Nguyen, Trang. 2008. "Information, role models and precived returns to education." *Mimeo.*

Obama, Barack. 2014. "State of the Union Address." Accessed 18.12.2015.
`http://www.whitehouse.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address`

Oreopoulos, Philip and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1):159–184.

Oreopoulos, Philip and Ryan Dunn. 2013. "Information and College Access: Evidence from a Randomized Field Experiment." *Scandinavian Journal of Economics* 115(1):3–26.

Papke, Leslie E. and Jeffrey M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11(6):619–632.

Papke, Leslie E. and Jeffrey M. Wooldridge. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics* 145(1-2):121–133.

Pinger, Pia R. 2015. "Understanding the Mechanisms behind Intergenerational Effects of Economic Distress." *Mimeo.*

Rotter, Julian B. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs* 80(1):1–28.

Stange, Kevin E. 2012. "An Empirical Investigation of the Option Value of College Enrollment." *American Economic Journal: Applied Economics* 4(1):49–84.

Statistisches Bundesamt. 2011. Die Neue Berufsbildungsstatistik. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2013. Berufsbildung auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2014. Schulen auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Stinebrickner, Ralph and Todd Stinebrickner. 2012. "Learning about academic ability and the college dropout decision." *Journal of Labor Economics* 30(4):707–748.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*a*. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies* 81(1):426–472.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*b*. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics* 32(3):601–644.

Taber, Christopher R. 2000. "Semiparametric identification and heterogeneity in discrete choice dynamic programming models." *Journal of Econometrics* 96(2):201–229.

Taber, Christopher R. 2001. "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies* 68(3):665–691.

Tolsma, Jochem, Ariana Need and Uulkje De Jong. 2010. "Explaining participation differentials in Dutch higher education: The impact of subjective success probabilities on level choice and field choice." *European Sociological Review* 26(2):235–252.

Trachter, Nicholas. 2015. "Stepping stone and option value in a model of postsecondary education." *Quantitative Economics* 6(1):223–256.

Turner, Sarah E. 2004. Going to College and Finishing College Explaining Different Educational Outcomes. In *College Decisions: How Students Actually Make Them and How They Could*, ed. Caroline Hoxby. University of Chicago Press for NBER.

Venti, Steven F. and David A. Wise. 1983. "Individual attributes and self-selection of higher education: College attendance versus college completion." *Journal of Public Economics* 21(1):1–32.

Wiswall, Matthew and Basit Zafar. 2015*a*. "Determinants of college major choice: Identification using an information experiment." *Review of Economic Studies* 82(2):791–824.

Wiswall, Matthew and Basit Zafar. 2015*b*. "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital* 9(2):117–169.

Wölfel, Oliver and Guido Heineck. 2012. "Parental risk attitudes and children's secondary school track choice." *Economics of Education Review* 31(5):727–743.

Zachary, Bleemer and Basit Zafar. 2015. "Intended College Attendance: Evidence from an Experiment on College Returns and Cost." *IZA DP No. 9445*.

Zafar, Basit. 2011*a*. "Can Subjective Expectations Data be used in Choice Models? Evidence on Cognitive Biases." *Journal of Applied Econometrics* 26(3):520–544.

Zafar, Basit. 2011*b*. "How Do College Students Form Expectations?" *Journal of Labor Economics* 29(2):301–348.

Tables and Figures

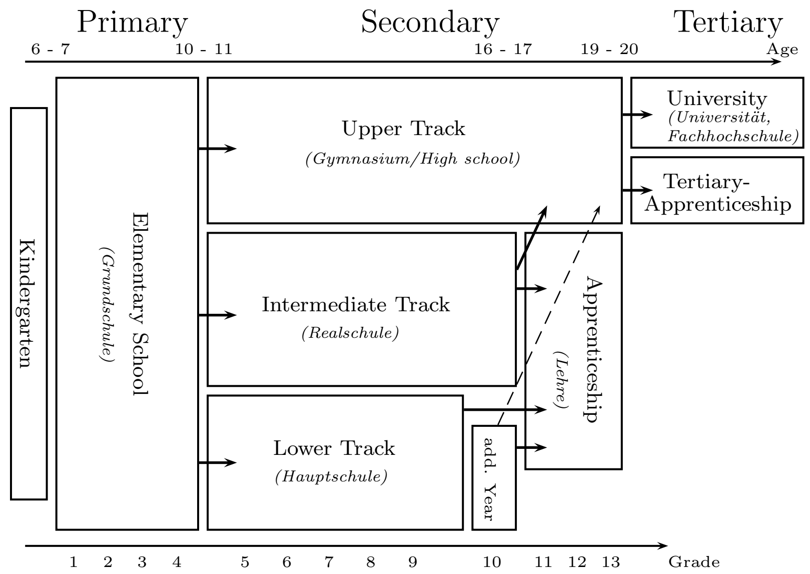
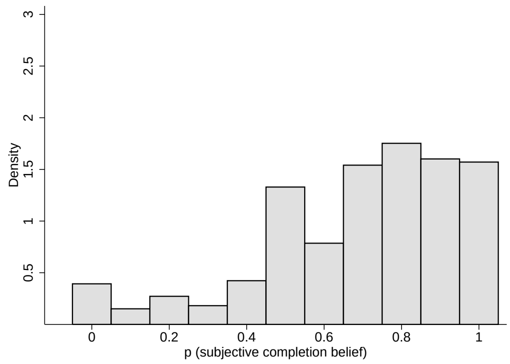
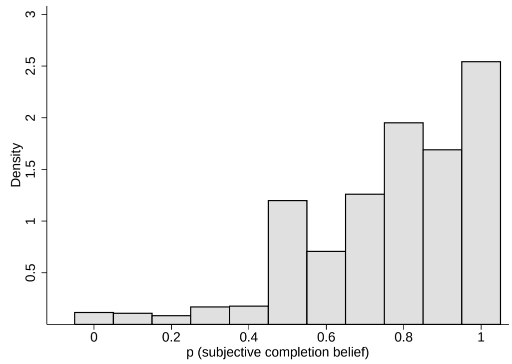


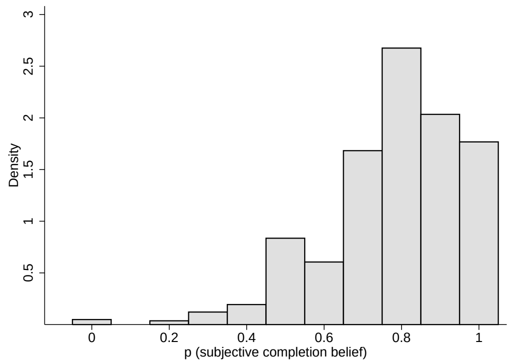
Figure 1.1: Simplification of the German education system
Source: Adaptation and extension of the overview provided by Wölfel and Heineck (2012).



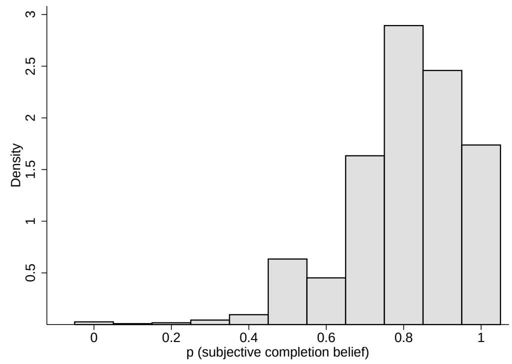
(A) No intentions (N=331)



(B) Apprenticeship (N=1'302)



(C) Tertiary apprenticeship (N=826)



(D) University (N=1'151)

Figure 1.2: Subjective completion beliefs by intentions-to-invest
Source: SOEP 2000-2013, own calculations.

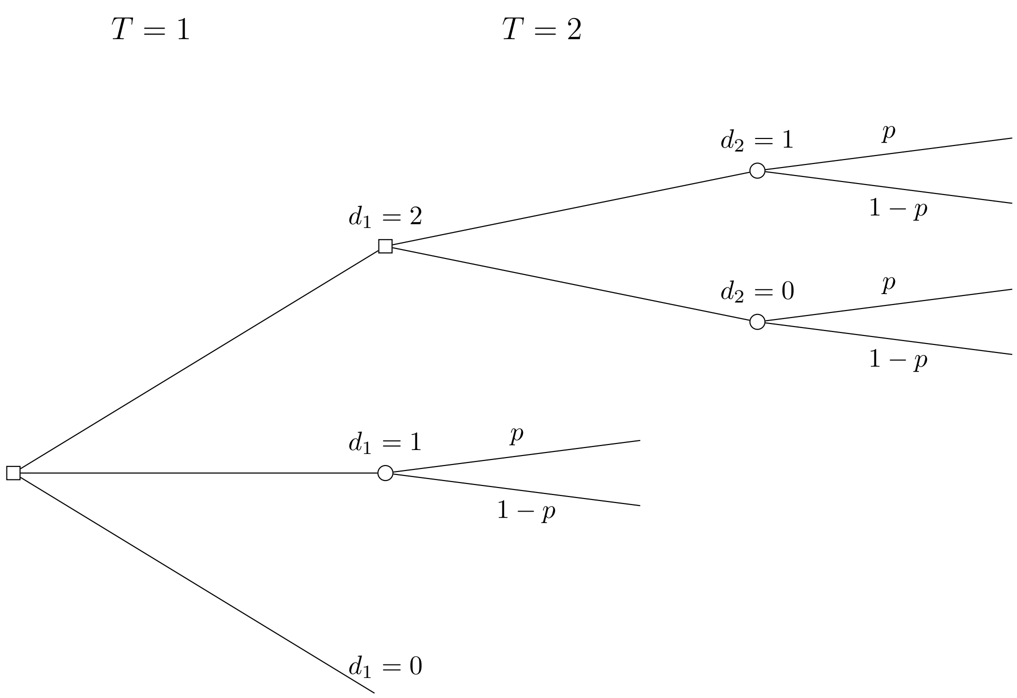


Figure 1.3: Sequential education decisions and timing of events

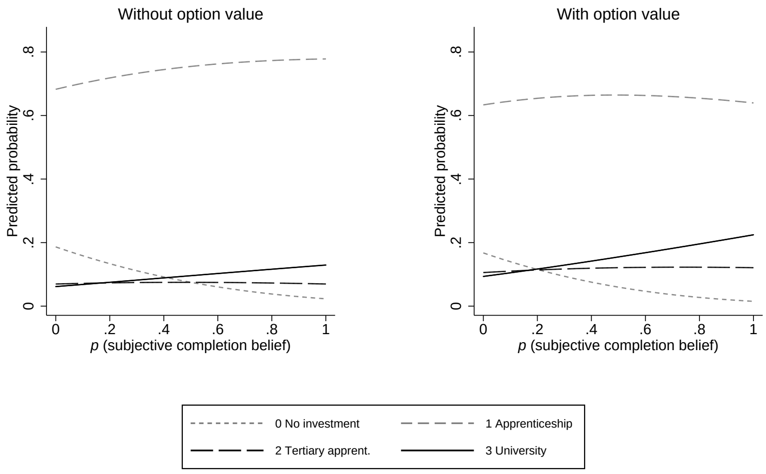


Figure 1.4: The role of the option value in a dynamic model of educational choice
Notes: Predicted probabilities constructed using estimates from Column (5) of Table 1.6 and evaluated at sample means.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

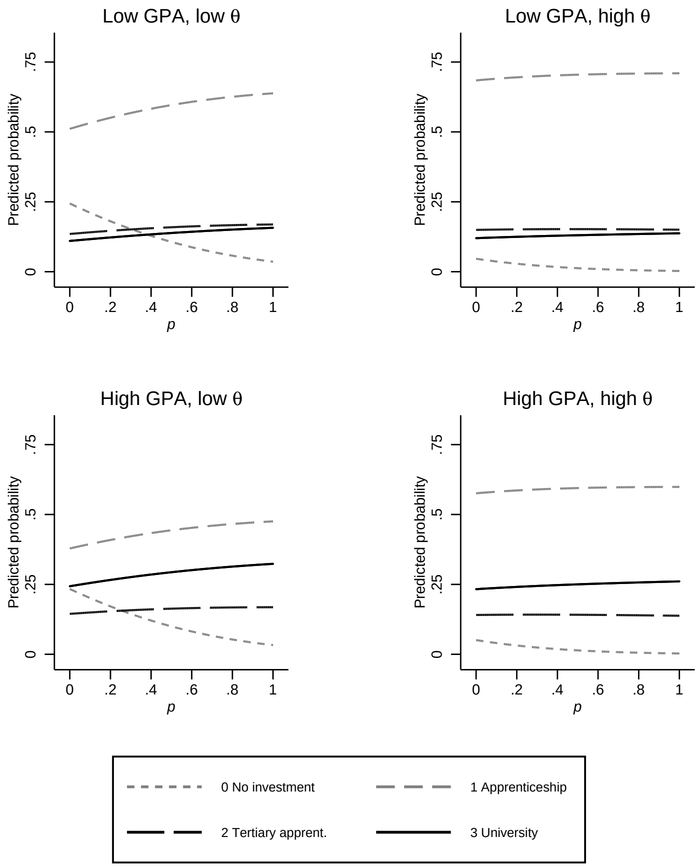


Figure 1.5: The role of academic ability and unobserved heterogeneity
Notes: Predicted probabilities constructed using estimates from Column (8) of Table 1.6, evaluated at sample means. High GPA = High $\theta = \Phi^{-1}(0.75)$, Low GPA = Low $\theta = \Phi^{-1}(0.25)$.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 1.1: Descriptive statistics by educational aspirations

	By aspiration level				Total
Variables	0	1	2	3	
p	0.692	0.768	0.782	0.805	0.776
	(0.258)	(0.220)	(0.176)	(0.155)	(0.198)
GPA (std)	-0.269	-0.288	-0.000	0.404	0.000
	(1.035)	(0.919)	(0.902)	(1.007)	(1.000)
Rec: Lower Track (yes/no)	0.166	0.265	0.044	0.034	0.132
	(0.373)	(0.441)	(0.204)	(0.181)	(0.338)
Rec: Intermediate Track (yes/no)	0.248	0.351	0.271	0.150	0.259
	(0.432)	(0.477)	(0.445)	(0.358)	(0.438)
Rec: High school (yes/no)	0.284	0.092	0.524	0.698	0.402
	(0.452)	(0.289)	(0.500)	(0.459)	(0.490)
In high school (yes/no)	0.272	0.015	0.596	0.767	0.411
	(0.446)	(0.123)	(0.491)	(0.423)	(0.492)
Locus of control (std)	-0.178	-0.169	0.082	0.195	0.003
	(1.014)	(0.963)	(0.861)	(0.862)	(0.928)
Risk attitudes (std)	-0.057	-0.006	-0.037	0.053	0.001
	(0.962)	(0.945)	(0.914)	(0.895)	(0.924)
Openness (std)	-0.178	-0.134	0.048	0.167	-0.000
	(0.990)	(0.956)	(0.879)	(0.926)	(0.942)
Agreeableness (std)	-0.103	-0.048	0.068	0.042	0.002
	(0.928)	(0.987)	(0.909)	(0.919)	(0.944)
Extraversion (std)	-0.135	-0.051	0.021	0.075	-0.002
	(0.948)	(0.914)	(0.919)	(0.989)	(0.945)
Neuroticism (std)	0.059	0.009	0.053	-0.066	-0.000
	(0.942)	(0.922)	(0.885)	(1.003)	(0.943)
Conscientiousness (std)	-0.132	0.053	-0.059	0.026	0.002
	(0.967)	(0.938)	(0.915)	(0.953)	(0.942)
Female (yes/no)	0.486	0.454	0.541	0.539	0.504
	(0.501)	(0.498)	(0.499)	(0.499)	(0.500)
Nr. of siblings	1.613	1.710	1.433	1.496	1.570
	(1.461)	(1.494)	(1.206)	(1.099)	(1.316)
Second-generation migrant (yes/no)	0.746	0.680	0.574	0.557	0.623
	(0.436)	(0.466)	(0.495)	(0.497)	(0.485)
Parent college-educated (yes/no)	0.199	0.101	0.306	0.495	0.283
	(0.400)	(0.301)	(0.461)	(0.500)	(0.450)
Parent cur. unemployed (yes/no)	0.124	0.160	0.087	0.045	0.103
	(0.330)	(0.367)	(0.282)	(0.208)	(0.304)
Log. net household income	10.019	9.890	10.624	10.855	10.377
	(2.231)	(2.216)	(1.358)	(1.295)	(1.834)
Cyclical youth unemployment (in Ror)	0.154	0.101	0.043	0.041	0.074
	(1.079)	(1.044)	(0.982)	(1.020)	(1.026)
Nr. of apprenticeship positions (in Ror)	98.380	98.544	98.538	99.368	98.791
	(4.906)	(5.261)	(5.600)	(5.124)	(5.279)
Nr. of students (in Ror)	23.700	22.711	24.156	25.730	24.095
	(14.204)	(14.354)	(13.991)	(14.091)	(14.223)
Nr. of high school graduates (in Ror)	26.081	25.755	27.289	27.758	26.775
	(6.313)	(6.526)	(6.218)	(7.064)	(6.673)
Nr. of Universities (in Ror)	10.789	9.620	10.916	11.381	10.585
	(10.304)	(9.666)	(10.037)	(9.988)	(9.938)
N	331 (9.17%)	1'302 (36.07%)	826 (22.88%)	1'151 (31.88%)	3'610

Note: Table presents sample means and standard deviations in brackets in total and by aspiration levels. Individual characteristics are assessed at the time of answering the Youth Questionnaire (with 17). GPA is the grade point average of German and Math grades, standardized and reversed, that higher values indicate better performance. Three indicators for school recommendations (with the age of 10), one indicator indicating if the student is currently in high school. Locus of control, openness, agreeableness, extraversion, neuroticism, and conscientiousness are principal components, std- stands for standardized to (0,1), where small deviations result from missings. We define second-generation migrants as having both parents born in a foreign country, parents college educated/unemployed if at least one has a college degree, is currently unemployed, cyclical component of youth unemployment in the region is extracted using the Hodrick-Prescott-Filter. The number of Universities in the region include all higher learning institutions.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 1.2: Determinants of subjective completion beliefs, OLS regressions

	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.029	0.029	0.029
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.025	0.026	0.025	0.028
	(0.014)	(0.014)	(0.014)	(0.015)
Rec: Intermediate Track (yes/no)	0.061	0.058	0.052	0.056
	(0.010)	(0.010)	(0.010)	(0.011)
Rec: High school (yes/no)	0.046	0.042	0.034	0.039
	(0.010)	(0.009)	(0.009)	(0.011)
In high school (yes/no)	0.005	0.003	-0.004	-0.005
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.022	0.019	0.019
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.008	0.005	0.005
		(0.004)	(0.004)	(0.004)
Openness (std)		0.005	0.004	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.008	0.008	0.008
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.016	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		0.001	0.002	0.001
		(0.004)	(0.004)	(0.004)
Conscientiousness (std)		0.031	0.034	0.034
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.013	-0.013
			(0.007)	(0.007)
Nr. siblings			-0.002	-0.003
			(0.002)	(0.002)
Second-generation migrant (yes/no)			-0.022	-0.010
			(0.007)	(0.013)
Parent college-educated (yes/no)			0.007	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			-0.002	0.001
			(0.013)	(0.013)
Log. net household income			0.007	0.007
			(0.002)	(0.002)
N	3,610	3,610	3,610	3,610
$\bar{p}$	0.776	0.776	0.776	0.776
$SD(p)$	0.198	0.198	0.198	0.198
adjR^2	0.052	0.107	0.115	0.116
Academic	+	+	+	+
F(pval)	30.470 (0.000)	20.679 (0.000)	17.564 (0.000)	17.029 (0.000)
Personality	-	+	+	+
F(pval)		24.168 (0.000)	24.626 (0.000)	24.328 (0.000)
Background	-	-	+	+
F(pval)			3.796 (0.000)	2.431 (0.013)
Labor market + FE	-	-	-	+
F(pval)				1.257 (0.184)

Note: Table presents coefficients, from linear regressions of subjective beliefs on varying sets of covariates, in (1) only on academic, (2) adds personality, (3) family background and individual characteristics, and (4) local labor market characteristics, region and time fixed effects (coefficients not presented). No recommendation is the base category, we include indicator variables for missing values in any of the covariates. Robust standard errors are given in brackets. We present the unconditional mean $\bar{p}$ and standard deviation $SD(p)$ of the dependent variable, the adjusted R^2, and joint significance tests. In the appendix we present analogous fractional response regressions.

Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 1.3: EFFECT OF SUBJECTIVE COMPLETION BELIEFS ON EDUCATIONAL OUTCOMES

	probit				probit eev			
					$\rho = .1$	$\rho = .3$	$\rho = .5$	$\rho = \hat{\rho}^o$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.921	0.809	0.726	0.716	0.612	0.383	0.120	0.561
	(0.142)	(0.146)	(0.148)	(0.150)	(0.149)	(0.141)	(0.129)	(0.168)
	[0.140]	[0.121]	[0.108]	[0.104]				
R^2_n	0.029	0.040	0.048	0.059				
$R^2_n(p)$	0.049	0.055	0.059	0.069				
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.776$, $SD(p) = 0.198$, $\hat{\rho}^o(se) = 0.147(0.003)$								
(B) Actual investment								
p	0.997	0.915	0.908	0.866	0.762	0.527	0.250	0.846
	(0.223)	(0.228)	(0.240)	(0.249)	(0.250)	(0.247)	(0.215)	(0.234)
	[0.069]	[0.062]	[0.056]	[0.044]				
R^2_n	0.087	0.100	0.122	0.184				
$R^2_n(p)$	0.113	0.121	0.141	0.199				
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.772$, $SD(p) = 0.201$, $\hat{\rho}^o(se) = 0.021(0.003)$								
(C) Actual investment, conditional on intentions								
p	0.901	0.845	0.836	0.726	0.622	0.393	0.129	0.716
	(0.256)	(0.262)	(0.272)	(0.276)	(0.277)	(0.269)	(0.240)	(0.141)
	[0.058]	[0.053]	[0.048]	[0.031]				
R^2_n	0.085	0.095	0.115	0.205				
$R^2_n(p)$	0.104	0.111	0.129	0.214				
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.781$, $SD(p) = 0.192$, $\hat{\rho}^o(se) = 0.010(0.002)$								
(D) Actual completion								
p	0.434	0.410	0.351	0.331	0.229	0.015	-0.214	0.326
	(0.181)	(0.185)	(0.189)	(0.192)	(0.190)	(0.190)	(0.171)	(0.094)
	[0.172]	[0.162]	[0.139]	[0.131]				
R^2_n	0.089	0.093	0.102	0.123				
$R^2_n(p)$	0.092	0.096	0.104	0.124				
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.769$, $SD(p) = 0.197$, $\hat{\rho}^o(se) = 0.005(0.003)$								
(E) Actual completion, conditional on intentions								
p	0.467	0.478	0.439	0.421	0.319	0.102	-0.135	0.418
	(0.198)	(0.202)	(0.206)	(0.210)	(0.210)	(0.191)	(0.197)	(0.083)
	[0.185]	[0.189]	[0.174]	[0.167]				
R^2_n	0.095	0.098	0.108	0.127				
$R^2_n(p)$	0.099	0.101	0.110	0.129				
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.778$, $SD(p) = 0.190$, $\hat{\rho}^o(se) = 0.004(0.003)$								
Academic	-	+	+	+	+	+	+	+
Personality	-	-	+	+	+	+	+	+
Family Background	-	-	-	+	+	+	+	+
Labor market	-	-	-	+	+	+	+	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. In the probit eev regressions we restrict the correlation between the errors to be 0.1, 0.3, 0.5 and to be equal to the selection-on-unobservables (the estimated is given by $\hat{\rho}^o$ along with its standard error). For each outcome in Panels (A) to (E), we present McFadden's pseudo-R^2 with and without p, and sample statistics for the varying subsamples. In the appendix we present analogous probit and bivariate probit regressions for dichotomized $p \geq 70\%$.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 1.4: DISAGGREGATED INTENTIONS-TO-INVEST, MULTINOMINAL PROBIT REGRESSIONS

	(1)	(2)	(3)	(4)
Apprenticeship, $d = 1$				
p	1.045	1.136	1.038	1.068
	(0.206)	(0.214)	(0.218)	(0.221)
	[0.021(0.029)]	[0.081(0.029)]	[0.081(0.030)]	[0.089(0.030)]
Tertiary Apprenticeship, $d = 2$				
p	1.094	0.907	0.802	0.776
	(0.210)	(0.221)	(0.226)	(0.230)
	[-0.011(0.034)]	[-0.003(0.035)]	[0.002(0.036)]	[0.001(0.036)]
University, $d = 3$				
p	1.557	1.073	0.881	0.827
	(0.217)	(0.227)	(0.233)	(0.237)
	[0.141(0.035)]	[0.052(0.035)]	[0.030(0.036)]	[0.021(0.035)]
N	3,610	3,610	3,610	3,610
LR(pval)	54.900(0.000)	36.009(0.000)	26.476(0.000)	26.237(0.000)
Academic	-	+	+	+
Personality	-	-	+	+
Family Background	-	-	-	+
Labor market	-	-	-	+

Note: Table presents, multinominal probit regressions of the educational intention-to-invest: drop out, apprenticeship, tertiary apprenticeship, and university on subjective beliefs and varying sets of covariates in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. Robust standard errors in round, average marginal effect along (with standard errors) in squared (round) brackets. The Likelihood Ratio (LR)-statistic measures the significance of p across equations.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 1.5: Disaggregated actual investment and completion, conditional on intentions, probit regressions

	Actual investment				Actual completion			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Apprenticeship ($d_1 = 1$)								
p	0.902	1.055	1.007	1.043	0.794	0.737	0.645	0.559
	(0.256)	(0.266)	(0.269)	(0.275)	(0.271)	(0.273)	(0.276)	(0.278)
	[0.133]	[0.151]	[0.136]	[0.090]	[0.292]	[0.271]	[0.237]	[0.204]
R^2_n	0.033	0.044	0.073	0.175	0.051	0.060	0.069	0.106
$R^2_n(p)$	0.054	0.071	0.096	0.196	0.064	0.071	0.077	0.111
Sample:	$N = 883$, $\bar{d}_1 = 0.948$, $\bar{p} = 0.769$, $SD(p) = 0.218$				$N = 502$, $\bar{d}_1 = 0.669$, $\bar{p} = 0.760$, $SD(p) = 0.220$			
(B) Tertiary apprenticeship ($d_2 = 1$)								
p	0.362	0.455	0.673	0.784	0.409	0.358	0.374	0.669
	(0.368)	(0.379)	(0.412)	(0.416)	(0.549)	(0.564)	(0.578)	(0.619)
	[0.120]	[0.150]	[0.219]	[0.250]	[0.098]	[0.089]	[0.094]	[0.129]
R^2_n	0.079	0.084	0.107	0.137	0.140	0.139	0.148	0.223
$R^2_n(p)$	0.081	0.087	0.112	0.143	0.142	0.141	0.150	0.226
Sample:	$N = 456$, $\bar{d}_2 = 0.965$, $\bar{p} = 0.781$, $SD(p) = 0.179$				$N = 314$, $\bar{d}_2 = 0.557$, $\bar{p} = 0.779$, $SD(p) = 0.177$			
(C) University ($d_3 = 1$)								
p	0.789	0.431	0.308	0.038	0.980	0.665	0.894	0.656
	(0.367)	(0.388)	(0.405)	(0.409)	(0.487)	(0.512)	(0.538)	(0.546)
	[0.282]	[0.153]	[0.108]	[0.013]	[0.253]	[0.172]	[0.226]	[0.155]
R^2_n	0.140	0.186	0.219	0.275	0.177	0.194	0.223	0.252
$R^2_n(p)$	0.146	0.188	0.220	0.275	0.186	0.198	0.229	0.255
Sample:	$N = 580$, $\bar{d}_3 = 0.978$, $\bar{p} = 0.801$, $SD(p) = 0.154$				$N = 428$, $\bar{d}_3 = 0.397$, $\bar{p} = 0.799$, $SD(p) = 0.156$			
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit regressions of investment (1)-(4) and completion (5)-(8) on subjective completion beliefs and varying sets of covariate, in (1/5) on in high school, region and time fixed effects, (2/6) adds academic, (3/7) adds personality, (4/8) family background, individual, and local labor market characteristics. We present McFadden's pseudo-R^2 and sample statistics for the varying subsamples. For some regressions the numbers of observations are slightly reduced.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table 1.6: Dynamic models of actual investment

	Dynamic model				Dyn. model with unobs. heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Apprenticeship ($d_1 = 1$)								
p	1.173	1.213	1.187	1.198	1.406	1.432	1.407	1.420
	(0.303)	(0.314)	(0.327)	(0.328)	(0.313)	(0.324)	(0.338)	(0.339)
θ					1.000	1.000	1.000	1.000
					(.)	(.)	(.)	(.)
High school ($d_1 = 2$)								
p	1.307	1.143	1.147	1.186	1.549	1.370	1.370	1.409
	(0.388)	(0.406)	(0.423)	(0.391)	(0.396)	(0.414)	(0.432)	(0.401)
θ					0.846	0.846	0.851	0.848
					(0.061)	(0.062)	(0.062)	(0.054)
University ($d_1 = 2, d_2 = 1$)								
p	0.462	0.105	0.063	0.084	0.462	0.103	0.062	0.082
	(0.259)	(0.273)	(0.282)	(0.301)	(0.260)	(0.273)	(0.282)	(0.300)
θ					-0.006	-0.006	-0.006	-0.007
					(0.047)	(0.047)	(0.047)	(0.049)
$\ln(\sigma_3)$				-0.719				-0.688
				(0.717)				(0.708)
N	2,116	2,116	2,116	2,116	2,116	2,116	2,116	2,116
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family+Labor market	-	-	-	+	-	-	-	+

Note: Table presents estimates of the model in equation (1.8). In the panel "Dynamic model", $\theta_i = 0$ for all i. The model in the panel "Dyn. model with unobs. heterogeneity" estimated by MSL with 100 random draws from $N(0, 1)$. The sets of covariates correspond to those in Table 1.3. Standard errors in parentheses. All regressions include an indicator for being in high school with 17, region and time fixed effects.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Chapter 2

An econometric model of health care demand with non-linear pricing

This chapter is jointly written with Rainer Winkelmann. A version of this paper is forthcoming in *Health Economics*.[1]

From 2004 to 2012, the German social health insurance levied a co-payment for the first doctor visit in a calendar quarter. We develop a new model for estimating the effect of such a co-payment on the individual number of visits per quarter. The model combines a one time increase in the otherwise constant hazard rate determining the timing of doctor visits with a difference-in-differences strategy to identify the reform effect. An extended version of the model accounts for a mismatch between reporting period and calendar quarter. Using data from the German Socio-Economic Panel, we do not find an effect of the co-payment on demand for doctor visits.

Keywords: Count data, Poisson process, co-payment, hurdle model
JEL classification: I10, C25

[1]We are grateful to two anonymous referees, Gregori Baetschmann, Helmut Farbmacher, Martin Grabka, Robert Jung, Joao Santos Silva, Kevin Staub, Joachim Winter, as well as participants of the 16th IZA/CEPR European Summer Symposium in Labour Economics for valuable comments, to Daniel Auer for very able research assistance, and to the German Institute for Economic Research (DIW Berlin) for providing access to the Socio-Economic Panel.

2.1 Introduction

Around 90% of the German population receive their health insurance coverage through the German statutory health insurance system (SHI). Before 2004, the SHI did not require any co-payment for doctor visits, although prescription drugs were subject to cost sharing for many years. Starting January 1st, 2004, the insured had to pay a 10 Euro fee for the first visit to a doctor in each calendar quarter ("Praxisgebühr"). Additional visits in the same quarter were free of charge. Thus the individual out-of-pocket expense became a non-linear function of utilization, dropping from 10 to 0 Euros after the first doctor visit in a quarter. Only individuals without any visit to a doctor could avoid paying the quarterly fee. Moreover, the fee did not apply to those with private health insurance (PHI). On January 1st, 2013, the co-payment was abolished.

Arguably, the introduction of a co-payment created an incentive to avoid doctor visits in a particular quarter of the year, at least at the extensive margin, for those close to indifference between consulting or not consulting a doctor. One therefore would expect that the probability of visiting a doctor at least once within a quarter should have fallen in the SHI population relative to the unaffected PHI population. By the same token, one might think that the number of doctor visits for those with at least one visit (the conditional-on-positives or intensive margin effect) should be unrelated to the co-payment (see Augurzky, Bauer and Schaffner, 2006, for such a view) although our analysis below will provide reasons why this is not necessarily the case.

A number of researchers have conducted quantitative evaluations of the effect of this reform on overall demand and demand at the extensive margin (Augurzky, Bauer and Schaffner, 2006; Schreyögg and Grabka, 2010; Farbmacher and Winter, 2013). So far, results have been mixed. Augurzky, Bauer and Schaffner (2006) report a negative and statistically significant difference-in-differences (DiD) coefficient in a logit model for "any visit". However, in their preferred specification they control for individual specific fixed effects, and the co-payment effect switches sign and becomes insignificant. Schreyögg and Grabka (2010) estimate a hurdle-at-zero negative binomial model and find no effect in either part of the model. Farbmacher and Winter (2013) find a statistically significant 4 percentage point reduction of the probability of any visit.

Our paper makes two main contributions. First, rather than using ad-hoc reduced form count data or binary response models to estimate the reform effect, we develop a new approach based on a structural model of health care demand. In our model, individuals are exposed to random health shocks arriving according to a homogeneous Poisson process. Individuals are myopic and decide at each instance whether or not to visit a doctor. In the random utility decision model, the out-of-pocket costs for seeing a doctor drop non-linearly after the first visit, leading to an increased hazard rate for subsequent visits. The model introduces a dynamic aspect, absent in econometric models used in prior work, where an increased cost of a first visit has two effects: The probability of no visit increases; moreover, a first visit, if it takes place, will tend to do so later in the quarter, lowering the overall number of subsequent visits. This does not require forward looking behavior of agents (which in fact is ruled out by our model).

The second contribution is an internally consistent approach to deal with a discrepancy between calendar quarter (i.e., pricing period) and reporting period. This is relevant if one uses survey data such as the Socio-Economic Panel (SOEP, Wagner, Frick and Schupp, 2007) where respondents are asked to state the number of visits during the three-months period preceding the day of interview. Often, the reporting period overlaps with two calendar quarters, and "reporting-mismatch" arises if, unrecorded in the data and thus unobserved by the analyst, a visit has taken place between the start of the relevant calendar quarter and that of the reporting period. In this case, the first visit in the reporting period has an effective price of zero under treatment, not of 10 Euros, as would be the case if the reporting period and the calendar quarter matched perfectly. Ignoring this issue leads to a misspecified likelihood function.

Farbmacher and Winter (2013) recommend to base estimation on the subset of individuals who were interviewed close to the end of a calendar quarter. Even if one can treat the interview day as random, this approach takes a heavy toll in terms of sample size and thus precision. For example, in our data, the number of available observations drops from 32,888 to 6,236 if the sample is restricted to those interviewed within a $\pm$ 10 days-window around the end of a quarter. On the other hand, using our assumptions on the underlying stochastic process, it becomes possible to derive the correct probability model for mismatched observations, and thus to employ all available data for estimation. Extensions to allow for unobserved heterogeneity are available.

Regarding data and identification, this paper largely follows the lead of the prior literature. For example, Augurzky, Bauer and Schaffner (2006) employ two waves from the SOEP, 2003 as pre-reform period and 2005 as post-reform period. The control group consists of people with private health insurance and provides a baseline counterfactual pre-and post reform trend in doctor visits. Any deviation from this baseline trend observed for the treated group (SHI) is then assumed to capture the effect of treatment. The co-payment was abolished in 2013, and we can use this as a second DiD experiment, by adding data on doctor visits from the 2012 and 2013 waves (for the subset of persons interviewed in the second quarter of 2013 or later). Assuming symmetric effects, a joint analysis of the two changes allows us to increase power.

While the main contribution of the paper is the development of a new model of demand for doctor visits with non-linear pricing, and thus methodological, our results add to the existing evidence regarding the lack of a robust effect of the 10-Euro co-payment on health care utilization. Perhaps, household data from the SOEP provide an indicator of utilization that is too noisy. Researchers might benefit from access to richer information provided by insurance level data, as in (Farbmacher et al., 2013). Or else, there really was no sizeable effect, because the amount was too small to be decision-relevant for most people, because it applied to the first visit only or because, in other cases, there just was no choice but to pay the fee, e.g., for the chronically ill.

2.2 Modeling the number of doctor visits

In this section, we derive the distribution of the individual number of visits Y_i during a fixed time interval $(0, T)$ representing a calendar quarter (i.e., $T = 90$ if time is measured in days). Suppose that sickness events arrive according to a Poisson process with constant rate λ_i. The total number of sickness events N_i during a quarter is then Poisson distributed with mean $\lambda_i T$. Let $X_{ij} \in \{0, 1\}$ denote the individual decision to visit a doctor ($X_{ij} = 1$) or not ($X_{ij} = 0$) at the j'th sickness event. The decision is made by comparing two utilities, utility u_{ij}^1 with a visit and utility u_{ij}^0 without. Utility depends on net income (after deducting direct cost c_j) and health: $u_{ij}^X = u((y_i - c_j)^X, h_{ij}^X)$, where it is understood that $c_j = 0$ when $X = 0$.

Assume linear utility $u_{ij}^1 = \alpha(y_i - c_j) + h_{ij}^1$ and $u_{ij}^0 = \alpha y_i + h_{ij}^0$. People choose to see a doctor if $u_{ij}^1 > u_{ij}^0$, i.e., $h_{ij}^1 - h_{ij}^0 > \alpha c_j$. The idiosyncratic health improvement from a visit, net of all non-pecuniary cost, has to be at least as large as the marginal utility of income times the pecuniary cost of a visit at sickness event j. Therefore, the probability of a visit is given by

$$\Pr(X_{ij} = 1) = 1 - \Pr(h^1_{ij} - h^0_{ij} < \alpha c_j) = p(c_j),$$

where $p(c_j)$ is decreasing in c_j as long as $\alpha > 0$.

With constant cost $c_j = c$, the probability of a visit is the same for all sickness events, and the total number of visits

$$Y_i = X_{i1} + \ldots + X_{iN} \tag{2.1}$$

has a Poisson distribution with mean $\lambda_i \times p(c)$ (Feller, 2008). Under the aforementioned reform, only the first visit during a quarter has a price of $c > 0$ whereas all subsequent visits are for free. Thus, the cost of a visit is path dependent. For example, the cost at the second sickness event can be written as $c_{i2} = c \times \mathbb{1}(X_{i1} = 1)$, and the probability of a visit at the second event is given by

$$\Pr(X_{i2} = 1) = p(c) \times p(0) + (1 - p(c)) \times p(c)$$

In general, it holds for all i that

$$\Pr(X_{i1} = 1) < \Pr(X_{ij} = 1) \qquad j = 2, \ldots, N$$

since $p(c) < p(0)$. With non-constant probabilities, (2.1) cannot be Poisson distributed.

To derive the distribution of Y_i in this case, it is useful to consider an alternative representation of the problem in terms of the underlying stochastic process. Specifically, Y_i is equal to the number of "renewals" (i.e. completed time spells between visits) during a fixed time interval. Inter-arrival times for a Poisson process are exponentially distributed. The non-linear pricing introduces a one-time jump in the hazard rate: $\lambda_{i0} = \lambda_i \times p(c)$ is the hazard rate for the time to first visit, and $\lambda_{i1} = \lambda_i \times p(0)$ that for the duration between subsequent visits. Under the assumptions of the model, $\lambda_{i0} < \lambda_{i1}$. This "non-stationarity" implies that the model does not correspond to a standard renewal process, and a new type of count data model is obtained.

2.2.1 The distribution of the number of visits

For simplicity of notation, we drop the "i" subscript in the following three subsections. Given the above assumptions, the time of the first visit t has an exponential distribution with rate λ_0, whereas the number of further visits during the quarter between t and T is Poisson distributed with rate λ_1, $Y(t,T) \sim Poisson(\lambda_1(T-t))$. Therefore, for $k \geq 1$, the total number of visits during a quarter has probability function

$$\Pr[Y(0,T)=k] = \int_0^T \frac{\exp(-\lambda_1(T-t))[\lambda_1(T-t)]^{k-1}}{(k-1)!}\lambda_0 \exp(-\lambda_0 t)dt \qquad . \tag{2.2}$$

where a first visit occurs between 0 and T, if at all, and we integrate over all these possible times. If we could observe t, we would directly estimate the parameters using the terms under the integral, $\Pr(Y=k-1|t,t<T;\lambda_1)$ and $f(t;\lambda_0)$. Our model applies to the case, where t is unobserved, as is typically the case in general purpose household or health surveys where just the number, and not the times, of visits is recorded.

Note that the time of first visit is not a choice variable in our model. It results from the interplay between a stochastic sickness arrival process (which is unaffected by the co-payment) and a utility maximizing choice that trades off the instantaneous health benefit of a visit with its monetary cost. One can show (see, e.g. Janardan, 1980; Baetschmann and Winkelmann, 2016) that the integral (2.2) has a closed form solution, and the probability function is given by

$$f(y;\lambda_0,\lambda_1) = \frac{\lambda_0\lambda_1^{y-1}\exp(-\lambda_0 T)}{(\lambda_1-\lambda_0)^y}\left[1-\sum_{j=0}^{y-1}\frac{\exp(-(\lambda_1 T-\lambda_0 T))(\lambda_1 T-\lambda_0 T)^j}{j!}\right] \qquad y=1,2,\ldots \tag{2.3}$$

and $f(0;\lambda_0,\lambda_1)=\exp(-\lambda_0 T)$. The model will be referred to as "dynamic hurdle" model, and we write $f(y;\lambda_0,\lambda_1) = DHurdle(y;\lambda_0,\lambda_1)$. If $\lambda_0=\lambda_1$, it can be shown that (2.3) simplifies to the probability function of the Poisson distribution.

2.2.2 Interpretation of parameters

The parameters of the model have a straightforward interpretation. λ_0 is the hazard rate for the first visit (or "stage 0" hazard), λ_1 the hazard rate for subsequent visits (or "stage 1" hazard). For instance, parameterizing $\lambda_0 = \exp(x'\beta_0)$, where x is a $(k \times 1)$ vector of covariates and β_0 a conformable vector of regression parameters, $\beta_0 \Delta x$ is the approximate relative change in λ_0 associated with a small change in x. In the context of two-part or zero-inflated models, β_0 and β_1 are often denoted as "extensive margin" and "intensive margin" effects, respectively. The mean of the model has generic form

$$\mathrm{E}(Y(0,T)) = \Pr(y > 0) + \mathrm{E}_t[\mathrm{E}Y(t,T)], \qquad 0 \leq t \leq T$$

Since $\mathrm{E}Y(t,T) = \lambda_1(T-t)$, where $T-t$ is the time from the first visit to the end of the calendar quarter, the intensive margin effect depends not only on λ_1 but on the expected duration of stage 1, and thus on λ_0, as well. In the model, a co-payment for the first visit means that it tends to happen later, leaving less time for accumulating further visits. It is therefore the case that the number of visits after the first visit is affected by the co-payment, even if λ_1 is not.

Using properties of the Poisson and exponential distributions, we obtain the following closed form expression for the mean:

$$\mathrm{E}(Y(0,T)) = \lambda_1 T + (1 - \lambda_1/\lambda_0)[1 - \exp(-\lambda_0 T)] \tag{2.4}$$

As required, the expected value of the distribution reduces to the Poisson mean when $\lambda_0 = \lambda_1$. The expected value is greater than λ_1 when $\lambda_0 > \lambda_1$, and smaller otherwise. A relative small value of λ_0 is an indication of "zero-inflation", or "extra-zeros", relative to the Poisson model, a situation encountered in many count data applications (Mullahy, 1986).

2.2.3 Discussion

The implied model for the first visit is identical to that used in a class of hurdle count data models introduced by Mullahy (1986). The probability function of the fixed hurdle model is given by

$$\Pr(Y=y) = \begin{cases} p_0(\lambda_0) & \text{for } y=0 \\ (1-p_0(\lambda_0))\dfrac{f(y;\lambda_1)}{1-f(0;\lambda_1)} & \text{for } y \geq 1 \end{cases}$$

where $f(y;\lambda_1)$ denotes the probability function of a standard count data model, e.g., Poisson or negative binomial distribution, and $p_0(\lambda_0)$ is a complementary log-log model. Pohlmeier and Ulrich (1995) argue that such a hurdle model can be appropriate for modelling the demand for health care. In their interpretation, the first contact decision for a general practitioner often triggers a number of re-appointments or referrals to specialists that are subject to a different mechanism and thus a different λ.

The standard hurdle model is not derived from an underlying stochastic process, however. It treats λ_0 and λ_1 as unrelated parameters that can be separately estimated. It ignores the random timing of the first visit, and thus the effect of λ_0 on the length of the period for which visits have a zero co-payment. In a fixed hurdle model, conditional-on-positives expressions such as $\Pr(Y=y|Y>0,x)$ or $\mathrm{E}(Y|Y>0,x)$ depend on λ_1 only, not on λ_0. It therefore rules out spill-over effects and cannot address path dependence generated by non-linear pricing.

By contrast, the dynamic hurdle model (2.3) naturally accounts for the timing of the first visit. The co-payment leads to a lower stage 0 rate and decreases the expected time available for subsequent visits. Although the timing of the first visit is unobserved, the corresponding count data model can be derived under the maintained assumptions. In the application below, we provide results for both fixed and dynamic hurdle models. While we argue that the dynamic model is *a-priori* preferable, it is of course possible that the underlying assumptions justifying the model are not satisfied in this particular application.

2.2.4 Identifying the effect of a co-payment on demand

In general, the two rates of the model can be expressed as functions of a number of exogenous factors x, such as prior health status, income, gender, employment status and the like. Suppose that $\lambda_{i0} = \exp(x_i'\beta_0)$ and $\lambda_{i1} = \exp(x_i'\beta_1)$. The above model suggests that with non-linear pricing, $\lambda_{i0} < \lambda_{i1}$, and thus

$$\exp(x_i'(\beta_0 - \beta_1)) < 1 \qquad \text{for all } i$$

However, attributing any such difference in rates to the existence of a co-payment for the first visit requires the absence of other explanations. But there are a number of factors that can rationalize a low initial rate and a higher one thereafter. Perhaps the leading explanation has been explored in the aforementioned paper by Pohlmeier and Ulrich (1995) argue, where visits occur in clusters and a first visit is followed by additional appointments for a given sickness spell. Thus a more convincing identification strategy uses difference-in-differences. Specifically, the co-payment did apply between 2004 and 2012 for those covered by SHI. Privately insured people were not affected and can serve as control group. We consider the following hazard rate specification:

$$\lambda_{it,j} = \exp(\theta_{t,j} + \beta_{1,j}\, SHI_i + \beta_{2,j}\, COPAY_{it} + x_{it}'\gamma_j) \qquad j = 0, 1, \qquad t = 2003, 2005, 2011, 2013$$

where $\theta_{t,j}$ are year dummies (2003 is dropped) and $COPAY_{it}$ is a dummy variable equal to one if the person is covered by SHI and the year is either 2005 or 2011, and else equal to zero. Thus, $COPAY_{it} = 1$ indicates active treatment, and $\beta_{2,0}$ is the extensive margin treatment effect under the "parallel trends assumption". This assumption implies that the counterfactual hazard rate for a first visit for the SHI population in the absence of a co-payment is equal to the actual SHI rate when no co-payment was in place (e.g., in 2003) multiplied by the appropriate trend growth factor obtained from the PHI population (e.g., $\exp(\theta_{2005,0})$).

Similarly, one could formulate a DiD model to estimate the effect on the (second) hazard rate for further visits, $\lambda_{it,1}$. This offers a kind of placebo test, as, within the above model, the reform did not change the incentives conditional on a first visit, and no effect should therefore be observed (i.e., the null hypothesis $H_0 : \beta_{2,1} = 0$ should not be rejected). There are good reasons not to put too much weight on such a test, though. First, there was a concurrent increase in the co-payments for prescription drugs on January 1, 2004, and we know from previous research that such out-of-pocket expenses tend to reduce the number of doctor visits as well (Winkelmann, 2004). Second, a referral from the first doctor was needed in order to receive free consultations by further doctors (or specialists). Thus, the co-payment may have increased the time- and effort costs of additional visits.

2.2.5 Dealing with mismatch

The empirical analysis is based on information from the SOEP on the number of visits "during the previous three months". Since interviews typically do not take place at the end of a calendar quarter, the reporting period overlaps with two calendar quarters. As noted by (Farbmacher and Winter, 2013) the standard models are invalid in this case. By contrast, our model prescribes a method to deal with mismatch in a theory consistent way, owing to the derivation of the dynamic hurdle model from an underlying stochastic process, which the standard hurdle model lacks. Consider a reporting period $(0 + r, T + r)$ that differs from the calendar quarter $(0, T)$ (e.g., $T = 90$ if time is measured in days). $r \in (0, T]$ is a known value in our data, since the day of the interview is recorded. Suppose a person gets interviewed on May 1st in a year. In this case, the relevant calendar quarter started on April 1st, $r = 30$, and the reporting period covers the final two months of the 2nd quarter and the first month of the 3rd quarter.

— — — Figure 2.1 about here — — —

This situation is illustrated in Figure 2.1, where visits are reported for periods B and C, whereas the calendar quarter includes A and B. In this case, the probability of no visit is the product of the probability of no visit in C times the probability of no visit in B, which in turn depends on whether or not a visit has taken place in A. In our model, the probability of a pre-reporting period visit is $\Pr(Y_A > 0) = 1 - \exp(-\lambda_0 r)$, and therefore,

$$\Pr(Y_B = 0) = (1 - \exp(-\lambda_0 r)) \exp(-\lambda_1 (T - r)) + \exp(-\lambda_0 r) \exp(-\lambda_0 (T - r)) \tag{2.5}$$

and

$$\Pr(Y_{B+C} = 0) = \Pr(Y_B = 0) \times \exp(-\lambda_0 r) \tag{2.6}$$

where $\Pr(Y_B = 0)$ is defined in (2.5). Ignoring mismatch would lead one to assume a different probability expression, in this case $\exp(-\lambda_0 T)$, and thus a misspecified model.

The expressions get more complex for $Y \geq 1$. The total number of events in the reporting period, Y, is then given by the sum $Y = Y_B + Y_C$, and

$$\Pr(Y = k) = \sum_{s=0}^{k} \Pr(Y_B = s) \Pr(Y_C = k - s) \tag{2.7}$$

This equation requires independence between two quarters which is guaranteed under the assumptions of the model. The assumption of full independence can be relaxed: independence conditional on observed or unobserved characteristics (e.g. a common log-normally distributed unobserved heterogeneity term) would be sufficient.

Equation (2.7) depends on two probabilities. The second probability is easy to establish since $\Pr(Y_C = k - s) = DHurdle(k - s; \lambda_0, \lambda_1, r)$. The first probability is a mixture of two distributions:

1. There has been at least one visit in A, i.e., the arrival time of the first event, t, predates r. In this case, counts in period B follow a Poisson distribution with mean $\lambda_1(T - r)$.

2. The arrival time of the first event exceeds r. In this case, counts in period B follow a dynamic hurdle model with parameters λ_0, λ_1, and $T - r$.

Combining terms,

$\Pr(Y_B = s) = (1 - \exp(-\lambda_0 r)) \times Poisson(s; \lambda_1(T - r)) + \exp(-\lambda_0 r) \times DHurdle(s; \lambda_0, \lambda_1, T - r)$

Substituting these expressions into (2.7), it is clear that

$$\Pr(Y_{A+B} = k) = \Pr[Y(0, T) = k] \neq \Pr(Y_{B+C} = k) = \Pr[Y(r, T + r) = k]$$

and the model is far from stationary (unless $\lambda_0 = \lambda_1$, of course). The key point is that the standard model assumes observation period and calendar quarter to be identical. If the two diverge, it is not clear whether the hazard of the observation period starts at λ_0 or at λ_1, since we do not know, whether or not an event has already taken place in the previous quarter.

As a corollary, all standard models used in the previous literature on first-visit co-payment effects (probit, logit and hurdle count models) are misspecified, and thus inconsistent when applied to a sample from survey data with mismatched reporting period. For consistent parameter estimation, one can use a subset of observations for which reporting period and calendar quarter are roughly aligned. But this only works, if the timing of the interview is random and not correlated with unobservable determinants of health utilization. It is therefore much better to estimate a model, such as the one derived here, that explicitly accounts for mismatch and is consistent even with non-random timing, and otherwise more efficient, by allowing for maximum likelihood estimation using the entire sample.

2.3 Data and Results

Data have been extracted from the Socio-Economic Panel (SOEP) that is made available by DIW Berlin. Four years are used, 2003, 2005, 2011 and 2013 and the analysis is restricted to individuals between the ages of 20 and 60. We do not impose a balanced panel, nor do we make the assumption of independence of observations across time. This affects the way standard errors should be computed, as we use clustered standard errors throughout. We consider two samples: The "restricted sample" includes all persons, who were interviewed within ± 10 days to the end of a calendar quarter. There are 6,236 such observations. In the restricted sample, the average distance to the nearest end-of-quarter is around 5 days. The second, the "full sample", includes all persons regardless of their time of interview. There are 32,888 such observations, and thus more than five times as many as in the restricted sample. The average distance to the nearest end-of-quarter in the full sample is about 24 days.

— — — Table 2.1 about here — — —

Table 2.1 reports means (and their standard errors) of variables employed in the estimation, separately for treatment group (SHI) and control group (PHI) and the two samples. Civil servants are excluded from the analysis due to their non-standard insurance arrangement. Selection into PHI is primarily based on income, as is evident in Table 2.1, where the mean log household income of PHI individuals is 0.5 above that of SHI individuals (in the full sample), corresponding to a 65 percent income difference. PHI individuals are also older on average (by about 3 years), more educated (by about 2 years) and more likely to be male. The regression analyses control for these factors, but the differences still raise the question of the comparability of the two groups, and hence of the validity of the parallel trends assumption underlying the DiD identification strategy. A formal test of this assumption was conducted by (Farbmacher and Winter, 2013), using the same kind of data from the SOEP, and they could not reject the null-hypothesis of parallel trends during the pre-treatment years.

The bottom panel of Table 2.1 splits the two samples further into two subsamples, the treatment years 2005 and 2011, and the non-treatment years 2003 and 2013 and shows for each group the average number of visits as well as the share with at least one visit. The non-treatment sample is somewhat smaller. The reason is that we had to drop all respondents who were surveyed during the first quarter of 2013, because their reporting period overlapped with the abolition of the co-payment by January 1st, 2013.

In terms of mean utilization, and using the full sample, we find that the SHI reported on average about 0.05 (or 2.5 percent) more visits in years without the co-payment in place than in years with co-payment. However, a similar trend is observed for the PHI individuals, who report about 0.1 fewer visits on average, so that the naive DiD effect, based on means only, is close to zero. A somewhat noisier picture emerges for the restricted sample, where the size of the control population (417 in treatment years and 310 in non-treatment years) leads to considerably more sampling uncertainty. If we look at the probability of any visit instead – the extensive margin – not much is found either. The probability actually increases somewhat for SHI individuals in years where the co-payment is in place. Of course, it would be premature to dismiss the possibility of a treatment effect based on this descriptive evidence alone.

2.3.1 Baseline result

In a next step, we estimated the fixed and dynamic hurdle Poisson models with a full set of control variables. The upper half of Table 2.2 shows the results. In terms of overall fit, the simple Poisson model is clearly inferior to the two-part generalizations that introduce different parameters for the utilization (yes/no) decision and for the intensity of use. The dynamic hurdle model leads to a substantially higher loglikelihood value relative to the fixed hurdle model (-12,690.4 as compared to -13,252.4).

Recall that λ_0 is the hazard rate for the time to a first visit. The probability of no visit is then equal to the survivor rate $\exp(-\lambda_0)$. Harzard rate and survivor rate are inversely related, i.e., factors increasing the hazard rate lower the probability of no visit, and vice versa. With the exponential parameterization, the displayed coefficients provide the predicted approximate relative change in the hazard rate associated with a unit change in the associated regressor. The exact relative change is obtained by applying the transformation $\exp(\hat{\beta}_j)-1$. For instance, according to the dynamic hurdle model, the hazard rate for a first visit for men is $(\exp(-0.357)-1)\times 100 = 30$ percent below that of women. Their predicted probability of no visit, evaluated at the mean probability of 0.64, is 10 percentage points above that of women.

— — — Table 2.2 about here — — —

There are some interesting asymmetries between first-visit hazard (λ_0) and that of subsequent visits (λ_1). For instance, income has no effect on the former, but a statistically significant negative effect on the latter, where a 10 percent increase in income is predicted to reduce the hazard rate for each further doctor visit by 1.5 percent. Taken at face value, this would mean that health care is an inferior good. More likely, the negative effect is due to a positive correlation between income and unobserved health status, capturing fewer visits by healthier individuals. As expected, individuals with disabilities have higher hazard rates in both states, and thus a higher predicted number of doctor visits, than individuals without.

The statistically insignificant point estimate of the treatment effect corresponds to a 1.6 percent reduction in the hazard rate for the first visit. The standard error is large, so sizeable positive or negative effects cannot be ruled out either. This finding is perhaps not too surprising, keeping in mind that a small change in the cost of a visit combined with a potentially low elasticity of demand should not have much of an effect. The effect on the stage 1 hazard (were a narrow interpretation of the model would predict none) is positive and relatively large, but again, the standard error is large and the coefficient is statistically insignificant as well.

2.3.2 Adding unobserved heterogeneity

All models can be extended to allow for unobserved heterogeneity, for instance by adding a multiplicative random effect to the exponential rates:

$$\tilde{\lambda}_{ij} = \exp(x_i'\beta_j)u_i$$

Here, u_i captures individual level differences in the demand for doctor visits, for example due to differences in latent health. In the context of our dynamic hurdle model, it seems reasonable to assume that the two rates for the first and for subsequent visits are multiplied by the same factor. Moreover, we will assume that the heterogeneity term is the same for repeated observations on the same individual, and thus $\tilde{\lambda}_{it,j} = \exp(x'_{it}\beta_j)u_i$.

The models discussed in Section 2 now hold conditional on u_i. To take the models to the data, the u_i term has to be eliminated from the likelihood function by taking expectations. Specifically, suppose that $\ln u_i$ is normally distributed, independently of x_i, with mean $-0.5\sigma^2$ and variance σ^2. Then u_i is lognormal with $\mathrm{E}(u_i) = 1$ and $\mathrm{Var}(u_i) = \exp(\sigma^2) - 1$. The marginal likelihood has no closed-form solution, but it can be numerically approximated using Gauss-Hermite quadrature.

The lower part of Table 2.2 provides estimation results for the three models with log-normal unobserved heterogeneity. In the case of the fixed hurdle model, heterogeneity is introduced only for the conditional-on-positives part, and the λ_0 estimates are thus identical to those in Table 2.2. Clearly, the improvements over the models without unobserved heterogeneity are large and statistically significant across the board. Unobserved heterogeneity changes the log-likelihood ordering of the three models, the fixed hurdle model having the highest log-likelihood in this case.

Point estimates for the important predictors of health care utilization, i.e., unemployment, disability and gender, are rather stable, regardless of whether unobserved heterogeneity is allowed for or not. The estimates for the reform effect remain statistically insignificant. Note that our specification imposes that the introduction and the abolition of the co-payment have the same effect size (and opposite sign), an assumption, that may be too restrictive. By 2012, many insured, in particular younger ones, had switched to new types of SHI contracts that offered co-payment waivers in return for joining a primary care model. We therefore estimated all models using the introduction sample only (2003 and 2005, results are available on request) but all substantive conclusions remain unchanged.

2.3.3 Full sample results

We also estimated the adjusted dynamic hurdle model, as discussed in Section 2.2.5, for the entire sample, in this case based on 32,888 observations. Otherwise, the specification remained unchanged. Results shown here are for the specification without unobserved heterogeneity. For the first visit, or stage 0, we obtained the log of the predicted hazard (with standard errors in parentheses):

$$\ln \hat{\lambda}_{it,0} = \underset{(.045)}{0.037}\,\mathrm{COPAY}_{it} + \underset{(.029)}{0.598}\,\mathrm{Disability}_{it} - \underset{(.017)}{0.353}\,\mathrm{Male}_i + \text{other terms}$$

The standard errors decrease roughly with the root of the sample size, compared the results presented in Table 2.2, and there is the expected efficiency gain from using the larger sample. The point estimates are in line with previous results. Again, no statistically significant reform effect is found.

For the stage 1 hazard, we get

$$\ln \hat{\lambda}_{it,1} = \underset{(.081)}{-0.041}\,\mathrm{COPAY}_{it} + \underset{(.040)}{0.537}\,\mathrm{Disability}_{it} - \underset{(.029)}{0.068}\,\mathrm{Male}_i + \text{other terms}$$

The finding of a large and about evenly distributed disability effect remains robust, as does the pattern that men have a substantially lower stage 0 hazard than women, whereas their stage 1 hazard does not differ much.

2.4 Concluding remarks

This paper introduced a new econometric model of health care demand under non-linear pricing, based on a Poisson process for the arrival of sickness events. In the model, a co-payment for the first visit during a calendar quarter potentially lowers the hazard rate for the first visit, leaving the subsequent hazard for further visits unchanged. The model was applied to an evaluation of a German health care reform of 2004 when a co-payment of 10 Euros was introduced for those covered by statutory health insurance. The co-payment was again abolished in 2013. In none of our various specifications, with or without unobserved heterogeneity and with or without reporting mismatch, did we find statistically significant effect of the co-payment on the number of doctor visits.

While the results thus are in line with those of two earlier studies by Augurzky, Bauer and Schaffner (2006) and Schreyögg and Grabka (2010) who also found no effect of the co-payment on utilization, the new methodological approach of this paper offers a number of additional insights that might prove useful for future research in related contexts. For instance, the perspective of a stochastic process is useful to understand that any changes to the first hazard likely also affects the distribution of additional visits, simply because it changes the time left in the quarter to accumulate such visits.

Second, the approach also points towards a theory consistent way to derive the likelihood of mismatched observations, i.e., observations for which reporting period and calendar quarter do not coincide. Such an approach avoids a loss of information incurred by limiting the sample to people interviewed at the end of a calendar quarter, and thus increases power. Of course, such data problems could also be addressed by using data from insurance claims rather than survey data. However, such claim data have their own problems. First, they usually include only a very limited set of socio-economic control variables, precluding certain types of analyses. Furthermore, they often do not allow for a DiD analysis, as it is unlikely that a useful control group can be established in a given claims dataset.

The models discussed in this paper should be useful in a number of other applications as well. For instance, in order to reduce absenteeism, firms have started to offer bonuses to workers with zero sick leave days per year. Depending on how the details of such schemes are designed, they may imply that the first absence is rather costly (the loss of the bonus) whereas subsequent absences have much lower cost. Similar non-linear pricing schemes can be observed for re-offenses in the context of fare dodging, were fines usually increase after the first offense. Moreover, this paper contributes to the literature on zero-inflated count data that are often encountered in health economic applications, adding a new, alternative model to the existing toolkit.

References

Augurzky, Boris, Thomas K. Bauer and Sandra Schaffner. 2006. "Copayments in the German Health System-Do They Work?" *RWI Essen: Discussion Papers 43.*

Baetschmann, Gregori and Rainer Winkelmann. 2016. "A dynamic hurdle model for zero-inflated count data." *Communications in Statistics - Theory and Methods, forthcoming.*

Farbmacher, Helmut and Joachim K. Winter. 2013. "Per-period co-payments and the demand for health care: evidence from survey and claims data." *Health Economics* 22(9):1111–1123.

Farbmacher, Helmut, Peter Ihle, Ingrid Schubert, Joachim K. Winter and Amelie A. Wuppermann. 2013. "Heterogeneous Effects of a Nonlinear Price Schedule for Outpatient Care." *CESifo Working Paper No. 4499.*

Feller, Willliam. 2008. *An introduction to probability theory and its applications.* 2 ed. John Wiley & Sons: New York

Janardan, Konanur G. 1980. "A stochastic model for the study of oviposition evolution of the pest callosobruchus maculatus on mung beans, Phaseolus aureus." *Mathematical Biosciences* 50(3-4):231–238.

Mullahy, John. 1986. "Specification and testing of some modified count data models." *Journal of Econometrics* 33(3):341–365.

Pohlmeier, Winfried and Volker Ulrich. 1995. "An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care." *Journal of Human Resources* 30(2):339–361.

Schreyögg, Jonas and Markus M. Grabka. 2010. "Copayments for ambulatory care in Germany: a natural experiment using a difference-in-difference approach." *European Journal of Health Economics* 11(3):331–341.

Wagner, Gert G., Joachim R. Frick and Jürgen Schupp. 2007. "The German Socio-Economic Panel Study (SOEP). Scope, Evolution and Enhancements." *Schmollers Jahrbuch* 127(1):139–169.

Winkelmann, Rainer. 2004. "Health care reform and the number of doctor visits-an econometric analysis." *Journal of Applied Econometrics* 19(4):455–472.

Tables and Figures

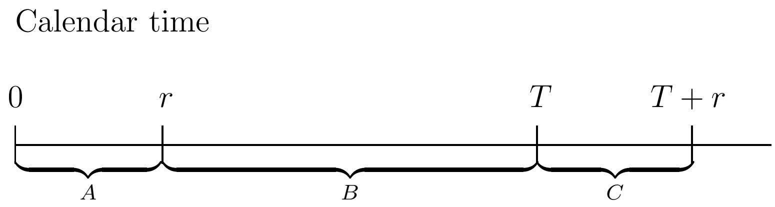


Figure 2.1: Mismatch between reporting period and calendar quarter

Table 2.1: Descriptive statistics by insurance status for restricted and full sample *(Means, cluster robust standard errors in parentheses)*

	Restricted Sample		Full Sample	
	PHI	SHI	PHI	SHI
Co-payment (yes/no)	0	0.580	0	0.577
		(0.006)		(0.002)
Distance to nearest end-of-quarter	4.642	5.254	23.721	23.811
	(0.115)	(0.043)	(0.224)	(0.079)
Age	44.281	40.899	43.890	41.004
	(0.418)	(0.162)	(0.226)	(0.090)
Male (yes/no)	0.597	0.452	0.614	0.447
	(0.021)	(0.008)	(0.012)	(0.004)
Years of schooling	14.096	12.032	14.148	11.969
	(0.124)	(0.038)	(0.071)	(0.022)
Disability (yes/no)	0.045	0.069	0.043	0.074
	(0.008)	(0.004)	(0.004)	(0.002)
Log net household income	10.954	10.490	10.962	10.466
	(0.025)	(0.009)	(0.014)	(0.005)
N	727	5,509	3,725	29,163
Years with co-payment (2005,2011)				
Number of doctor visits	1.830	2.048	1.902	2.072
	(0.135)	(0.051)	(0.067)	(0.024)
Any visit (yes/no)	0.607	0.643	0.603	0.652
	(0.025)	(0.009)	(0.011)	(0.004)
N	417	3,197	2,096	16,833
Years without co-payment (2003,2013)				
Number of doctor visits	2.216	2.039	1.988	2.123
	(0.197)	(0.061)	(0.076)	(0.028)
Any visit (yes/no)	0.619	0.638	0.622	0.649
	(0.027)	(0.010)	(0.012)	(0.004)
N	310	2,312	1,629	12,330

Note: PHI: private health insurance; SHI: social health insurance.

Source: Socio-Economic Panel (SOEP), version 30, doi:10.5684/soep.v30

Table 2.2: Poisson and hurdle models of health care utilization, restricted sample (N=6,236)

	Poisson	Fixed Hurdle		Dynamic Hurdle	
	λ	λ_0	λ_1	λ_0	λ_1
Panel A: without unobserved heterogeneity					
SHI	-0.196	0.007	-0.234	0.079	-0.447
	(0.094)	(0.082)	(0.093)	(0.076)	(0.169)
Co-payment (yes/no)	0.158	0.041	0.150	-0.016	0.297
	(0.119)	(0.104)	(0.123)	(0.098)	(0.209)
Unemployment (yes/no)	0.329	-0.014	0.369	-0.079	0.567
	(0.080)	(0.078)	(0.076)	(0.068)	(0.122)
Disability (yes/no)	0.823	0.817	0.535	0.654	0.563
	(0.055)	(0.075)	(0.054)	(0.060)	(0.086)
Male (yes/no)	-0.330	-0.380	-0.146	-0.357	-0.105
	(0.038)	(0.036)	(0.039)	(0.033)	(0.061)
Log net household income	-0.105	-0.053	-0.091	-0.025	-0.145
	(0.032)	(0.030)	(0.033)	(0.028)	(0.054)
Log likelihood	-14,781.7	-13,252.4		-12,690.4	
Panel B: with unobserved heterogeneity					
SHI	-0.263	0.007	-0.241	0.104	-0.363
	(0.105)	(0.082)	(0.095)	(0.103)	(0.155)
Co-payment (yes/no)	0.275	0.041	0.175	-0.016	0.244
	(0.134)	(0.104)	(0.138)	(0.131)	(0.191)
Unemployment (yes/no)	0.288	-0.014	0.271	-0.119	0.542
	(0.093)	(0.078)	(0.095)	(0.093)	(0.117)
Disability (yes/no)	0.772	0.817	0.613	0.794	0.584
	(0.069)	(0.075)	(0.089)	(0.080)	(0.079)
Male (yes/no)	-0.335	-0.380	-0.185	-0.421	-0.150
	(0.038)	(0.036)	(0.045)	(0.044)	(0.056)
Log net household income	-0.090	-0.053	-0.106	-0.043	-0.111
	(0.032)	(0.030)	(0.032)	(0.037)	(0.049)
$\ln(\sigma)$	-0.002	-0.296		-0.517	
	(0.020)	(0.024)		(0.023)	
Log likelihood	-11,976.5	-11,781.3		-11,833.6	

Note: Dependent variable: Number of doctor visits. All models include three year dummies, a constant, a quadratic in age and level of schooling. Standard errors in parentheses are clustered at the individual level. PHI: private health insurance; SHI: social health insurance.

Source: Socio-Economic Panel (SOEP), version 30, doi:10.5684/soep.v30

Chapter 3

Analyzing educational achievement differences between second-generation immigrants: Comparing Germany and German-speaking Switzerland

A version of this chapter is published as: Kunz, Johannes S. 2016. "Analyzing Educational Achievement Differences between Second-Generation Immigrants: Comparing Germany and German-Speaking Switzerland." *German Economic Review* 17(1): 61-91.[1]

In this study, I provide evidence that the educational achievement of second-generation immigrants in German-speaking Switzerland is greater than in Germany. The impact of the first-generation immigrants' destination decision on their offspring's educational achievement seems to be much more important than has been recognized by the existing literature. I identify the test score gap between these students that cannot be explained by differences in individual and family characteristics. Moreover, I show how this gap evolves over the test score distribution and how the least favorably-endowed students fare. My results suggest that the educational system of Switzerland, relative to the German system, enhances the performance of immigrants' children substantially. This disparity is largest when conditioning on the language spoken at home, and prevails even when comparing only students whose parents migrated from the same country of origin.

Keywords: Immigrant comparison; Educational achievement decomposition; Germany and Switzerland;
JEL classification: I21; I24; J15;

[1] I thank the editor, Michael Lechner, two anonymous referees, Gregori Baetschmann, Christian Dustmann, Winfried Pohlmeier, Philippe Ruh, Ruben Seiberlich, Florian Schaffner, Kevin Staub, Raphael Studer, Steven Stillman, Rainer Winkelmann, and seminar participants at Luzern, Zermatt, Zurich, the 'Education and Equality of Opportunity' Workshop at the ZEW, Mannheim, and the 'Junior Economist Workshop on Migration Research' at the ifoCEMIR/CESifo institute, Munich for very helpful comments and suggestions, and Daniel Auer for very helpful research assistance.

3.1 Introduction

This study contributes to a growing body of literature evaluating the educational performance of children born to first-generation immigrants in Western European countries (*inter alia*, Algan et al., 2010; Belzil and Poinas, 2010; Dustmann, Frattini and Lanzara, 2012; Heath, Rothon and Kilpi, 2008; Lüdemann and Schwerdt, 2012; Schneeweis, 2011; Song, 2011). The literature has documented severe disadvantages faced by second-generation immigrants in terms of educational achievement, wage income, and unemployment probabilities relative to their host countries' native peers. In addition to these relative assessments, I will argue that absolute achievement and learning processes of second-generation immigrants need to receive greater attention. Focusing on second-generation immigrant students alone helps to answer the question of what the parental sorting decision implies for the educational opportunities of their children. Likewise, it indicates the effectiveness of the host countries' educational institutions in accommodating the needs of immigrants' children. Understanding their absolute learning process can help policy makers to turn their immigrant populations into a productive strength in society.

The economics of education literature has concentrated on integration by assessing within-country educational differences between children of natives and first-generation immigrants. These relative within-country differences are then compared across countries. Conditioning on individual and family background, native-immigrant achievement gaps are reduced significantly but remain at high levels in most Western European countries. However, this approach cannot reveal the relevant parameters of the immigrant children's learning process due to the following reasons: First, imagine a family deciding upon a destination country, the relevant counterfactual is what would be the educational achievement of their offspring had they decided for another country, not had they been native parents in the chosen destination. Second, from the reduction in native-immigrant achievement differences alone it cannot be concluded that the performance of these immigrants' children is satisfactory. Instead the reduction of the performance gap might be entirely due to conditional changes in performance of natives' children. Moreover, when conditioning on the educational background of the parent population, these educational backgrounds have to be comparable across heterogeneous home countries to be meaningful. It is possible, if not probable, that having received secondary education in Turkey captures a different proficiency level than having completed the same education in Germany. Additionally, the covariate cells for native students with parents that have *no primary education* are empty or nearly empty in most developed countries. The variable *language spoken at home* – which has recently gained prominence in native-immigrant comparisons – is most problematic in this regard; we do not know to whom we compare the immigrants' children, and what these conditional differences tell us. Third, little is known about which educational institutions support the absolute performance of second-generation immigrants in Western European countries. Assessing language acquisition using a non-integration based approach seems to be most fruitful, since this learning process differs between immigrants' and natives' children.

Using the Programme for International Student Assessment (PISA) 2009 survey, I compare second-generation immigrant students' reading test scores directly across countries. Large scale and internationally comparable performance tests like PISA facilitate comparisons of students' educational achievement across countries. A drawback of these large-scale student evaluations in assessing second-generation immigrants is missing information on pre-migration characteristics of the parent generation, such as the reason for or the time of migration. Self-selection and self-sorting of migrants to host countries create heterogeneous immigrant populations across countries. To account for this selectivity without observing pre-migration characteristics, I focus on Germany and the German-speaking part of Switzerland. Comparing these two regions has several advantages in dealing with potential self-selection and self-sorting of first-generation immigrants: First, both countries are high immigration countries that have experienced a similar migration history, resulting in relatively homogeneous immigrant populations (Castles, 1986), which allows for an assessment of the country of origin. This is important when the human capital of the parents differs by their country of origin. Second, because they are in a similar language area, the results will be confounded neither by language differences nor by self-selection of immigrants into a certain language environment. Third, comparing countries with the same test language allows for a meaningful assessment of differences in reading literacy. Being a measure of language acquisition, it is highly relevant for the immigrants' children assimilation and learning processes.

The key contribution of this study is the comparison of reading literacy between immigrants' children in Germany and German-speaking Switzerland. First, I show that second-generation immigrant-native gaps diminish in both countries when conditioning on parental background characteristics, as commonly found in previous studies. Next, I decompose the achievement difference between second-generation immigrant students in Germany and German-speaking Switzerland into a component attributable to differences in background characteristics and a component that cannot be explained by those characteristics.[2] The decomposition into explained and unexplained components is performed parametrically and semi-parametrically to allow for non-linear impacts of the background characteristics and failures in the out-of-support validity. Then, I show how the unexplained gap evolves over the test score distribution, and provide evidence separately for unfavorably endowed children of immigrants, Turkish descendants the largest overlapping immigrant population, and native students for reasons of comparability. Finally, I present how the gap varies with school characteristics that might support the immigrants' children learning process.

The results suggest that the performance of immigrants' children in Switzerland is substantially higher than in Germany. This disparity is largest for very low-performing and unfavorably endowed second-generation immigrants. Differences reduce but prevail when conditioning on the parents' country of origin and when restricting attention to Turkish descendants. By contrast, the improvement in educational achievement does not extend to children of native-born parents, as they score almost as well in Germany as in German-speaking Switzerland. Among several school characteristics one that seems to explain a large part of the disparity between immigrants' children is the average test score performance of pupils in school.

The remainder of this article proceeds as follows: In the next section, I discuss the literature, historical patterns of migration into Germany and Switzerland, and their educational systems; Section 3.3 describes the PISA 2009 dataset, covariates used, and the econometric procedure; Section 3.4 presents the results and a suggestive discussion on possible reasons for the difference in educational achievement; and Section 3.5 concludes.[3]

[2]Due to the missing information on pre-migration characteristics, the resulting differences are interpreted as decompositions rather than causal effects. Another reason for the decomposition interpretation is that there is no obvious manipulable policy action, as discussed by Fortin, Lemieux and Firpo (2011).

[3]There are three appendices in the Supplementary Material: Appendix C.1 replicates the analysis using math literacy as the outcome variable; Appendix C.2 presents robustness checks for all language areas of Switzerland, all immigrants in both countries, sampling weights, different matching estimators, plausible values, and different imputation procedures; and Appendix C.3 presents information about the sample selection, missing values/imputation, common support, and covariate balance before and after matching.

3.2 Literature, migration history, and school systems

First, I highlight the present article's approach in comparison to those pursued in the literature. Subsequently, I briefly summarize the migration policies and histories of the two countries to motivate the sample of comparison, and discuss their educational systems.

3.2.1 Previous literature

The economic literature on educational achievement of second-generation immigrants in Western Europe – based on internationally comparable performance tests – focuses predominantly on integration. In these studies, integration is taken to be the difference in test score performances of immigrants' children and their host countries' native peers; cross-country studies then compare these national gaps across countries. The most detailed study has been performed by Dustmann, Frattini and Lanzara (2012). In their study, the PISA 2006 survey is used to analyze test score disparities between second-generation immigrant and native students across a large number of countries. They find that within-country gaps reduce substantially when taking into account the intergenerational correlation, as proxied by parental education levels. For Germany and Switzerland, their analysis reveals that even after conditioning on the children's family background, the gap in Germany is double the size of the gap in Switzerland.

In this study, I compare these second-generation immigrant students directly, instead of comparing their within-country immigrant-native test score gaps. This comparison has two advantages: First, within-country immigrant-native comparisons are problematic, since it is not clear whether differences originate from differential performance among the children of natives, immigrants, or both. Moreover, comparing the educational achievement of natives to that of second-generation immigrants, even when conditioning on parental education, might be misleading. For example, if receiving primary education in Turkey is different – in terms of knowledge acquisition – from primary education in Germany or Switzerland and if the intergenerational correlation in education operates through knowledge transmission, then conditioning on parental education will not reveal the effect of the destination country's school system upon immigrant children. Additionally, some covariate cells are empty and out-of-support validity is at least dubious. For example, in industrialized countries the category *no primary education* for native parents is almost empty as a result of compulsory schooling laws. A related problem arises when conditioning on the *language spoken at home*, a procedure that traditionally reduces the second-generation immigrant-native gap significantly. By this conditioning it is not clear to whom we compare these immigrant students, or what the conditional correlations tell us.

Second, related but different is the question whether educational researchers and policy makers shall concentrate on relative or absolute performance. In other words, how to weight the educational *equity-efficiency trade-off* when focusing on immigrants' children. Here, I do not aim to take a general stand on this subject. Generally, each comparison is interesting in its own right. Comparisons to natives' children can, for example, answer questions about discrimination, which the literature has documented in some detail. On the other hand, comparing immigrants with immigrants (in different countries) can identify which institutions support the immigrant students' learning process, in particular, when their learning process is different from the one of natives' children as in (second-)language acquisition. Complementing the literature by this non-integration based measure is the central motivation of this study.

So far, little is known about which educational institutions promote the second-generation immigrants' learning process, in an absolute sense, irrespective of their native peers. A notable exception is Levels, Dronkers and Kraaykamp (2008), who compare second-generation immigrants' performance directly. By pooling destination countries, their study precludes a detailed comparison across individual countries. Schnepf (2008) studies countries separately by comparing educational inequalities using the dispersion between the 5th and the 95th percentile of the test score distribution within countries. She finds that both Switzerland and Germany have a 10-20% larger dispersion within the performance of second-generation immigrants than within natives. Schnepf (2008) argues that liberal migration policies in Western Europe created heterogeneous populations of first-generation immigrants within countries, which led to substantial inequality among their children through intergenerational transmission. Her study highlights the importance of accounting for the heterogeneity among immigrants' children within countries by looking at the distribution of test scores.

This study also fits into a developing branch of the literature that introduces new reference groups in order to assess selectivity or origin effects. Dronkers and Heus (2010) investigate negative selection of immigrants by studying their difference from non-emigrant peers. Dustmann, Frattini and Lanzara (2012, 170) address "the opportunities or disadvantages migration implies for the children of immigrants" by comparing children of Turkish immigrants to their peers who have not emigrated, and thus were born and raised in Turkey. Country of origin effects are assessed within a single destination country (e.g., Luthra, 2010; Ours and Veenman, 2003) or a single country of origin in different destination countries, again relative to the host countries' native peers (e.g. Song, 2011). My study adds to this new branch of the literature by including those whose parents have emigrated to different destination countries. This is an important reference group, because it reveals the consequences implied by the parental sorting decision. Moreover, this study includes a larger number of countries as sources of immigration, as well as a cross-country dimension.

3.2.2 Migration history

When comparing second-generation immigrant students across countries it is essential to find suitable comparison groups. Unfortunately, internationally comparable student assessments do not contain information on pre-migration characteristics of the foreign-born parents and across countries it is hard to find overlapping immigrant populations, especially when considering the country of origin. Therefore, I compare Germany and German-speaking Switzerland as they attracted very similar immigrant populations due to their migration policy regimes and language environments. In the following, I briefly summarize the relevant migration histories until 1994 when the second-generation immigrants are born (within the countries of testing).[4]

After the Second World War, war losses and post-war reconstruction lead to a substantial under-supply of un- and semi-skilled labor in Western Europe. The employment-to-population ratio was further diminished by low birth rates, extended compulsory education, and increasing life expectancies. Industrial expansion and new methods of mass production created an extensive demand for labor migration into Western Europe (Castles, 1986).

In 1948, Switzerland established large-scale imports of labor based on bilateral agreements with Italy, followed by Germany in 1955 (Liebig, 2004). According to Hansen (2003), recruitment in Southern Europe was due to the expectation of a smoother assimilation into the labor market compared to more distant areas or ethnicities. Both countries then started to recruit

[4]This section summarizes the overview of Castles (1986) and draws from Schmid (1983), Liebig (2004), and Zimmermann (1995).

in Spain and Greece – in reaction to increasing competition for cheap labor and exhaustion of Southern European labor resources – they turned to Turkey, Morocco, Portugal, Tunisia, and Yugoslavia.

In Switzerland, employers recruited for themselves, but admission and organization was centralized by the Swiss government. The German government created a state recruitment administration, controlled by the Federal Labor Office. Employers had to apply for foreign labor and the Federal Labor Office set up recruitment offices in Mediterranean countries to select suitable workers. Complex legal and administrative frameworks were put in place to regulate and control foreign labor, aiming to prevent settlement by maintaining rapid turnover, a common feature of all European guest-worker programs (Castles, 1986).

By the sixties, international competition and employers' requests for a more stable workforce induced the governments of Switzerland and Germany to liberalize foreign labor policies. This initiated the phase of family migration, which allowed workers to reunify with their families. In 1963, Switzerland introduced a ceiling on the stock of foreigners per firm, which was rather unrestrictive and therefore replaced by *global quotas* in 1970 (Liebig, 2004, 164). These quotas set an upper limit to newly entering labor migrants into the country. In the wake of the oil crisis in 1973, the guest-worker systems came to a halt in all European countries.

In the 1980s, the ban of recruitment left family migration and later the asylum migration as the only channels to legally enter the German or Swiss labor markets. Conversely to the expected return migration, only a few of the former guest-workers returned to their home countries.[5] Most had settled and could not be expelled. Asylum migration became substantial after 1989. First, the fall of the Iron Curtain led to large inflows of Eastern Europeans. Second, the Balkan war pushed many Yugoslavian refugees and asylum-seekers into Western Europe (Hansen, 2003; Algan et al., 2010).

In 1991 both countries reorganized their labor migration. In Germany, nationals of countries that were not part of the European Economic Community or some other exceptions were only allowed to fill vacancies in sectors with unmet labor demand. The Swiss government introduced the *Three-Circles-Model*. The first cycle granted preferential status for nationals from the European Economic Area, in the second cycle immigrants from the United States, Canada,

[5]In 1983, Germany offered financial incentives for voluntary repatriation, but only a few immigrants responded to the policy.

Australia, and New Zealand could be recruited if demand could not be met within the first cycle, and the third cycle included nationals from all other countries who could only be recruited on a subsidiary basis (Liebig, 2004).

Overall, these patterns of migration were similar in both countries and resulted in homogenous immigrant populations (at least relative to other country pairs).[6] Nevertheless, despite the similarities, there are some notable differences in the immigrant populations of Germany and German-speaking Switzerland: First, the reintegration of ethnic Germans – called *Aussiedler* – from Poland and the Former Soviet Union is only observed in Germany. Second, Western European (from Germany, France, or Lichtenstein) and Albanian immigrants are only observed in Switzerland. These groups could possibly have had other reasons to migrate than guest-workers and their relatives. Hence, I exclude these three categories as they have no equivalent in the other country.[7] Students descending from other countries overlap. Although there are more immigrants descending from former Yugoslavia in Switzerland than in Germany, and the reverse pattern for immigrants from Turkey, their compositions will be balanced by the estimation procedure explained below.

Restricting the comparison countries to the same language area accounts for several selection aspects, such as language preferences. Yet, it is important to keep in mind, that the every-day spoken language in German-speaking Switzerland is *Swiss German* (a variety of dialects) which is not fully equivalent to *Standard German*. Nevertheless, in school children learn the written language *Swiss Standard German*, which is similar in most respects to *Standard German*. This should favor the children of immigrants in Germany, since they are exposed to *Standard German* not only in school but also in every-day spoken language.

3.2.3 Educational systems

In both counties, the educational systems are decentrally governed and organized by federal states: 16 "Bundesländer" in Germany and 26 "Kantone" in Switzerland. PISA assesses 15-year olds, hence participants of the 2009 wave were born between 1993 and 1994 in the respective

[6]In the literature, it is common to contrast the German or Swiss experience with countries that have very different immigration policies such as traditional countries of migration like Canada or Australia (Entorf and Minoiu, 2005) or the United Kingdom and France (Algan et al., 2010).

[7]The sample proportions before and after exclusion can be found in Appendix C.3. I present the main results including these three groups of immigrants in Appendix C.2, Table C.2.2; the effects are similar to the preferred specification.

country of testing. In this section, I briefly summarize the main features of the educational systems within this time period. Table 3.1 presents some key indicators of the school systems based on the sample of immigrant students used throughout the study (and explained in Section 3.3.1).

— — — Table 3.1 about here — — —

Within the children's first three years of life parents had optional access to early childhood care in both countries. From age three to six, children can visit a type of preschool called *Kindergarten*. In the PISA 2009 survey 92.38% of the immigrants' children in Germany and 98.77% in German-speaking Switzerland report that they have attended at least one year of Kindergarten. In Table 1, the average time spent in preschool is slightly longer in Switzerland. Despite the focus on second-language acquisition due to the different dialects in Switzerland, there were no country-wide institutional arrangements on how to support non-German-speaking children of immigrants. Every Kanton developed its own institutions of which most offered courses in German as a second-language already in Kindergarten (EDK, 2002). Similarly, there was no unified approach to support immigrants' children in Germany. Governed by the federal states, attention was paid on the testing of German language skills before entering primary school. Children identified with poor language skills were offered courses in German as a second language (KB, 2006).

In both countries, compulsory primary education starts with the sixth birthday with cut-off dates ranging from 30^{th} of June to 30^{th} of September in Germany and, with rare exceptions, form 30^{th} April to 30^{th} June in Switzerland. Hence, the effective range of school entrance ages lies between late five years and early seven years of age. In the sample, the average age at school entry is 6.40 in Germany and 6.69 in German-speaking Switzerland.

Tracking is generally organized similar and takes place early on in the children's school career. In Germany, tracking occurs after 4 years of school (with the exceptions of Berlin and Brandenburg that track in sixth grade). Generally in Switzerland tracking takes place later, after 5 to 6 years (and rare exceptions in fourth grade). In both countries, tracking is mainly based on teacher recommendations and grades in primary school. In both countries, there is some evidence for discrimination between immigrants' and natives' children at the transition from primary to secondary school (see Lüdemann and Schwerdt (2012) for Germany and Häberlin, Imdorf and Kronig (2004) for Switzerland).

At the time of testing, when the children are at the age of 15 (or early 16), the majority of second-generation immigrant children is attending grade nine (56.47% in Germany and 66.87% in Switzerland). The average grade level of 15-year olds is 8.97 in Germany, and 8.80 in German-speaking Switzerland. Both countries have fairly high rates of grade repetition. The probability of repeating one or more grades is 34% in Germany and 30% in German-speaking Switzerland (where the probability for repeating more than one grade is small, approximately 4% in Germany and 1% in Switzerland).

In the discussion of the results below, I address some aspects of the school systems that might cause differential performance of immigrants' children in the two countries. In particular, I consider the amount of German lessons (per week), as well as the overall amount of lessons (per week), the proportion of second-generation immigrants in school, and the average performance among fellow immigrants' and natives' children in school.

It is important to note that there might be other explanations instead of the educational systems that might cause a disparity in performance between the two countries, such as attitudes towards immigrants or integration efforts in general. Yet, Mayda (2006) suggests that attitudes towards immigrants are similar and Liebig (2004) argues for parallels in integration efforts. Still, children of immigrants may perceive their inclusion into the host society differently and expect for example greater returns to education in Switzerland than in Germany. This might create incentives to invest in education and knowledge acquisition that in turn result in higher test scores. Another explanation could be that the Swiss educational system simply better fits the test. Yet, PISA evaluates "skills for life" that capture what is considered to be necessary knowledge independently of the student's curricula and that appear to be of particular importance for students with a migration background who need at the very least be able to actively participate in their host societies. In addition, it could be that intrinsic motivation to perform well on a test is different between the two groups (e.g., Segal, 2012). However, this too can be considered as an important skill that is relevant for later performance in life. If it is also resulting from the new environment it could be argued that it should be part of the achievement difference.

3.3 Data, estimation strategy, and interpretation

In this section, I present the PISA 2009 survey, the sample selection process, and the background characteristics. Subsequently, I describe the parametric decomposition developed by Blinder (1973) - Oaxaca (1973) [henceforth BO] and the semi-parametric propensity score matching decomposition.

3.3.1 Data

A comprehensive summary of the PISA dataset is given in OECD (2009); here, I briefly summarize the features relevant for my analysis. PISA is an internationally standardized achievement test with mean (of 500) and standard deviation (of 100), facilitating an interpretation in terms of percentage points of the international standard deviation. The target population is 15-year olds enrolled in school. PISA evaluates the students' "knowledge and skills for life" in three categories: Reading, math, and science literacy. I concentrate the discussion on reading literacy results, since I believe that reading literacy and language acquisition are integral parts of the immigrants' assimilation and learning processes.[8] Due to difference in every-day spoken language – German compared to Swiss German – the results might differ depending on the competences considered. I therefore included the results based on math literacy test scores in Appendix A of the Supplementary Material. The results are qualitatively the same, though the differences are larger in magnitude and of greater statistical significance.

Similar to Dustmann, Frattini and Lanzara (2012), I define second-generation immigrants as being born in the country of testing while having both parents born in a foreign country. This definition excludes children with one foreign and one native-born parent, as they are found to be statistically different from children that have both parents born in a foreign country (Ohinata and van Ours, 2012).

Ammermüller, Heijke and Wößmann (2005) note that missing information on students' background characteristics mainly stem from low-performing students, and is thus not missing at random but shall be imputed. Accordingly, I perform median imputations on the school level (including native students in school) of the variables *mother* and *father education*, the *highest*

[8]Moreover, reading literacy was the central focus of the PISA 2009 survey with most testing time allocated. In each wave, the central focus of PISA changes. It was science literacy in 2006 and math literacy in 2003.

occupation status, and *number of books at home*.[9] One observation with missing information on *gender* has been dropped. If the *language spoken at home* is missing, I coded it as being different than the national testing language. As discussed above, I dropped the children of immigrants whose parents both originated from Western Europe, *Aussiedler*-countries or Albania.[10] Children with a mixed foreign background – with both parents born in a foreign country but from different areas – are included in the *another origin* category. Since the children of immigrants form a selective subpopulation of the overall student population in both countries, sampling weights are not likely to recover the target population of interest. The sampling design is the same in both countries, hence selection is unlikely to be correlated with the country indicator. For that reason, I refrain from using sampling and replication weights in the main analysis.[11]

This selection generates a sample size of 1,180 second-generation immigrant students: 824 in Switzerland and 356 in Germany.[12] In Table 3.2, the descriptive statistics of the background characteristics are presented.

— — — Table 3.2 about here — — —

In the top row, the average reading literacy test scores exhibit already a substantial difference between the countries. The reading literacy is a standardized test measuring reading comprehension. PISA provides five plausible values, of which I take the average.[13] Overall, the characteristics seem to be relatively similar in both countries.

[9]In Appendix C.1, Table C.3.1 and C.3.2 show that the missing values are positively correlated with each other and negatively with the test scores and that dropping these will change the outcome considerably. The total number of observations with at least one imputed value is 473. The number of imputations/missings due to the covariates can be found in Table C.3.3. In order to show that my results are not driven by the imputation mechanism, I present in Appendix C.2 Table C.2.7 a median imputations based on country level and in B.8 a regression based imputation that takes into account the covariance structure of the imputed covariates. The results are very similar. For the school characteristics presented in Table 3.1, I first impute the values only on immigrant students in school (to account for example for extra German lessons), and if these are not sufficient I impute them including natives.

[10]There is more information about the sample selection in Appendix C.1 Table C.3.4. The results including these groups of immigrants' children can be found in Appendix C.2 Table C.2.2.

[11]The main specification using sampling weights is presented in Appendix C.2 Table C.2.3. Using replication weights instead of bootstrapping seems to result in smaller standard errors (results not reported).

[12]Hereafter, the Swiss second-generation immigrant population always refers to those in the German-speaking part. As a sensitivity check, I estimate the decompositions including the French- and the Italian-speaking parts which exhibit a similar pattern (cf. Appendix C.2, Table C.2.1). The Swiss sample includes the PISA extension survey for cantonal representativeness and has therefore a larger number of observations.

[13]In Appendix C.2 Table C.2.6, I present the main results using only one plausible value as recommended by the PISA manual, results are almost indistinguishable.

There are some differences in the educational levels of the parent generation as measured by the International Standard Classification of Education (ISCED), which is assessed by four dummy variables the first capturing no formal education, the second primary up to lower secondary, the third measures upper secondary/non-tertiary, and the fourth theoretically oriented tertiary education. In Switzerland there is a smaller share of uneducated parents and a larger share of parents who have only completed primary or lower secondary education. Conversely, in Germany a larger share of the parent generation obtained an upper secondary degree, and the proportion completed a tertiary education is smaller than in Switzerland.

On the other hand, the *highest occupational status* measured by the Highest Socio-Economic Index of Occupational Status (HISEI) – that ranks occupations by the returns to education and takes the highest one among the parents – is almost identical in both countries.[14] There is a considerably larger number of immigrant families that speak a language other than German at home in Switzerland (81%) than in Germany (67%). In conventional immigrant-native comparisons these students are necessarily out-of-support. Finally, as discussed above, in the Swiss sample more children of Southern European immigrants (mainly Italians), less Turkish, and more former Yugoslavians are represented. As previous studies indicated, there is correlation in the test score performance of immigrants' children and their descent (e.g., Dustmann, Frattini and Lanzara, 2012; Dronkers and Heus, 2010; Song, 2011). Despite of the disparity between the proportions, the overlap of types of immigrants in Switzerland and Germany is much greater compared to other country pairs that have been contrasted in the literature. This enables me to control for the country of origin in greater detail than previous studies.

3.3.2 Estimation strategy

The goal of this paper is to compare the average reading test scores of immigrants' children in Switzerland $\bar{Y}_{CHE}$ and Germany $\bar{Y}_{DEU}$

$$\Delta = \bar{Y}_{CHE} - \bar{Y}_{DEU} \tag{3.1}$$

and to decompose this average test score gap Δ, into a part that can be associated with the covariates described above and the remaining part that is not attributable to these background

[14]For more information on this index, see Ganzeboom, De Graaf and Treiman (1992).

characteristics. The latter can capture, for example, greater integration into the host society, or more inclusive school institutions.[15] These two effects are decomposed by simulating the mean and the distribution of individual and family backgrounds of the students in Switzerland (Germany) within the distribution of students' background characteristics in Germany (Switzerland). In other words, reweighing the student population in the one country to reproduce the covariate-distribution of the other.

In the parametric BO decomposition, the covariate-adjusted mean is estimated by performing separate linear regressions of test scores on characteristics for both groups and combining the estimated coefficients of one regression with the covariate vector of the other regression (Blinder, 1973; Oaxaca, 1973). Adding and subtracting this estimated covariate-adjusted mean, Equation (3.1) can be written as

$$\Delta = \underbrace{\hat{\beta}_{CHE}(\bar{X}_{CHE} - \bar{X}_{DEU})}_{\Delta_X} + \underbrace{(\hat{\beta}_{CHE} - \hat{\beta}_{DEU})\bar{X}_{DEU}}_{\Delta_S}, \tag{3.2}$$

or equivalently in the reverse decomposition

$$\Delta = \underbrace{(\hat{\beta}_{CHE} - \hat{\beta}_{DEU})\bar{X}_{CHE}}_{\Delta_S} + \underbrace{\hat{\beta}_{DEU}(\bar{X}_{CHE} - \bar{X}_{DEU})}_{\Delta_X}, , \tag{3.3}$$

where Δ_X refers to the difference due to characteristics, and the main interest lies in Δ_S which presents the difference not explained by characteristics, called the *structure effect*.[16] The covariate vector X includes different sets of explanatory variables: I use *Other* covariates as a baseline specification which includes *gender*, *age in months*, *educational level of parents* (four dummies each), *highest occupation of the parents*, and *number of books at home* (six dummies). Additionally, I control for *German spoken at home* (one dummy variable) and the *country of origin* (four dummies), separately and jointly.

Matching generalizes the BO decomposition such that it does not rely on assumptions regarding functional form or out-of-support validity (Ñopo, 2008). It accounts for the possibility that the background characteristics have non-linear impacts, and that conditioning on several covariates might create subcategories that have no equivalent in the other country. Moreover,

[15]A similar cross-country decomposition strategy was taken by Ammermüller (2007). He decomposes the PISA test score gap between Germany and Finland, although, not specific to immigrants' children.

[16]For a comprehensive treatment of decomposition methods see Fortin, Lemieux and Firpo (2011).

matching decomposition can be performed on the propensity score without imposing additional assumptions (Frölich, 2007).

The matching estimator replacing, for example, $\hat{\beta}_{CHE}\bar{X}_{DEU}$ is the kernel-weighted average over the test score distribution of second-generation immigrants in Switzerland:

$$\frac{1}{N_{DEU}} \sum_{i \in I_{DEU}} \sum_{j \in I_{CHE}} w(i,j) Y_{j,CHE}$$

where $w(i,j)$ are the kernel weights that weigh observations according to the similarity of their propensity scores (background characteristics) to those of the other countries' students. N_{DEU} (N_{CHE}) is the number of immigrants' children in Germany (Switzerland), and I_{DEU} (I_{CHE}) is the set of immigrants' children in the common support of the other country. Analogous to the parametric procedure, the covariate-adjusted mean is added and subtracted from Equation (3.1) to write it as a sum of the components Δ_X and Δ_S (see Frölich, 2007, for a more comprehensive treatment).

Propensity scores are estimated by logit regressions of a country dummy on the same characteristics as in the parametric decomposition. The supports overlap greatly and imposing the common support therefore discards only very few observations.[17] The kernel weights are constructed by a Nadaraya-Watson kernel regression with Gaussian kernel and bandwidth held constant at 0.1 across specifications.[18] The quantile gaps are calculated by the horizontal differences between estimated quantiles of the actual and the covariate-adjusted test score distributions, constructed by the matching specification. Standard errors are bootstrapped in all decompositions with 500 replications, and the propensity scores are re-estimated in each replication.

3.3.3 Interpretation

In principle, any unexplained gap between Germany and Switzerland can be due to differences between the countries (e.g. school institutions) or due to unobserved differences in the composition of the first-generation immigrant populations. Under the assumption that there is no selection bias conditional on the included background characteristics, the gap represents the

[17] In Appendix C.1 Figure C.3.1 I present the distribution of propensity scores over the common support.

[18] I present the main results using Nearest Neighbor matching with 1 and 5 neighbors and bandwidths 0.95 and 0.105 in Appendix C.2: Table C.2.4 and C.2.5. The effects are similar to those in my preferred specification.

genuine country effect. In consequence, the observed test score performance of the matched immigrants' children in one country identifies the counterfactual outcome, i.e. the performance of the immigrants' children had their parents migrated to the respective other country. In the following, I discuss possible threats to the validity of this assumption. The assumption would be violated, if unobserved variables differ between the countries and correlate with children's test score performance.

Despite of the similar migration histories and recruitment efforts, differential self-selection of migrants between Switzerland and Germany might violate the assumption. For example, it could be the case that more motivated individuals decided to emigrate to Switzerland (which might not be accounted for by the included covariates). If their motivation is transmitted to their children and subsequently translated into higher test scores, then the unexplained part in the decomposition would comprise this selection bias. Indeed, the parental education appears to be slightly more favorably distributed among immigrants in Switzerland (cf. Table 3.2). This could imply a positive bias in the unexplained part in favor of Switzerland. Yet, the number of books at home – a control variable intended to capture parents' esteem in education and academic success (Schütz, Ursprung and Wößmann, 2008) – and the parents' occupations are almost indistinguishable between the two countries. Moreover, there are reasons to belief that the scope for such a selection bias is limited. First, the guest-worker scheme allowed migrant workers to temporarily leave their home countries to work and accumulate savings before they returned to their home countries. Schmid (1983) argues that this was in line with the intentions of the migrant workers. Yet, they were unable to accumulate sufficient savings and found themselves trapped in their host countries where they became permanent migrants (see also Castles, 1986). This unplanned migration pattern probably prevented sophisticated migration decisions. Second, it seems unlikely that they gathered sufficient information that enabled them to differentiate in detail between the two countries, given how similar the countries must have seemed to an outsider. Third, the migration costs must have been very similar to migrants from the same area, because Germany and Switzerland shared a common language, geographic location and prospering economic condition.

In addition, the parents might differ by the country they received their education in, which cannot be ruled out due to missing pre-migration information. For instance, it could be the case that the parents that migrated to Germany had acquired some of their education in Germany while those who migrated to Switzerland migrated after they finished their education in their home countries. However, I expect that most of these differences are controlled for by conditioning on the parents' level of education, occupation, and especially whether they speak German at home. All of these factors should correlate strongly with the country where an individual received its education in. Furthermore, as discussed in detail above the similar migration patterns suggest similar life-cycle stages of the migrants, i.e. guest-workers must have finished their education before migrating.

Another potential confounder is the return migration. Even when both sending populations would have been identical, if the migrants that returned to their home countries differed between the countries then the disparity will entail a selection bias due to return migration. Nevertheless, as discussed above, the return migration was minor in both countries; conversely to the attempts of the respective governments (see, *inter alia*, Castles, 1986; Liebig, 2004).

Finally, it could be the case that the differences in test scores stem from the differences in the ethnic composition of the immigrant populations (rather than from composition of the country of origin which is controlled for in the analysis). However, so far little is known about how the educational performance differs by ethnicity (of the second-generation) or if the composition of ethnicities differs between the two countries in a significant manner. It seems very promising to assess potential differences resulting from differences in ethnicity. Unfortunately, large scale and internationally comparable student assessments do not contain information on ethnicity which prevents a detailed analysis of ethnicity.

In sum, I focus the comparison on immigrant populations which migrated from similar areas to similar countries that share the "same" language and migration history. Moreover, I balance out differences by controlling for important background characteristics such as the parents' education level, their occupation, the language they speak at home, and their origin. Still, it is inherently possible that there is some selection based on unobservables. However, the analogies between the Swiss and the German migration experiences are rarely observed across other country pairs and time periods. Keeping these concerns in mind, it is interesting to answer how much of the test score disparity, observed in Table 3.2, can be explained by the covariates described above. Moreover, comparing the performance of those immigrants' children can shed light on the learning process of immigrants' children and what the parental migration decision implied for their children.

3.4 Results

In this section, I provide answers to the question "How would the children of immigrants perform in Switzerland (Germany) if they had the same distribution of background characteristics as those in Germany (Switzerland)?" I start by replicating the commonly used within-country second-generation immigrant-native test score regressions, in order to provide a reference for my main results.

— — — Table 3.3 about here — — —

Table 3.3 presents separate regressions of reading test scores on an immigrant indicator and the covariate vectors explained in Section 3.2 for Germany and Switzerland. The results are comparable to those in Dustmann, Frattini and Lanzara (2012, Table 3.4) for the survey of 2006. In Column 1 Panel A, the unconditional reading test score gap between second-generation immigrant and native students is -58.32 test score points in Germany and -48.72 in Switzerland (Panel B). Including individual and family characteristics reduces the gap significantly, as presented in Column 2. Furthermore, the gap narrows substantially by adding the *German spoken at home* indicator, leaving only -7.91 points remaining unexplained in Germany and -6.62 in Switzerland (Column 3). In Columns 4 to 6, I present the same procedure for the subsample of immigrants exposed to a German-speaking environment and with the exclusion of *Aussiedler*, Albanians, and Western European immigrants' children. The regression coefficients are larger in Germany, implying that *Aussiedler* have driven the gap downwards. In Switzerland, the selected subsample seems to perform slightly worse than the overall sample. As discussed in detail earlier, it is impossible to draw conclusions regarding which country better supports the educational achievement of second-generation immigrants from these results alone.

Turning to my main analysis, Figure 3.1 depicts the unconditional reading test score densities of immigrants' children in Germany and in German-speaking Switzerland (left panel).

— — — Figure 3.1 about here — — —

The left graph shows that reading test scores tend to be higher in Switzerland. The right graph depicts the unconditional quantile gap, which is the horizontal difference between the countries' distribution functions at various quantiles. The quantile gap is positive almost everywhere and is particularly large and statistically significant among low-performing students. This indicates, that the low-performing second-generation immigrants score substantially higher in German-speaking Switzerland than in Germany. In the following, I present evidence that this relationship holds when conditioning on background characteristics.

3.4.1 Mean difference

I start by decomposing the average reading test score difference, shown in Table 3.4. In Panel A, I adjust the second-generation immigrant population in Switzerland to match the characteristics of the second-generation immigrants in Germany. In Panel B, I reversely adjust the children of immigrants' characteristics in Germany to match those in Switzerland.

Column 1 presents the unconditional average reading test scores of immigrants' children in Switzerland (457.09) and Germany (439.05). The unconditional mean difference Δ_S is 18.05, which is both significantly different from zero and large in magnitude (and of course equivalent in Panel A/B). As a reference, this is roughly 30% of what an additional school year adds in my subsample of immigrants' children.[19]

Columns 2-5 present the parametric BO decomposition and Columns 6-9 the semi-parametric matching decomposition. Column 2 (and 6) uses the vector of *Other* background characteristics as above, 3 (and 7) adds the *German spoken at home* indicator, 4 (and 8) controls instead for the *country of origin*, and the last specification in Column 5 (and 9) uses in addition both language and origin indicators.

— — — Table 3.4 about here — — —

Starting with Panel A Column 2, the covariate-adjusted mean among Swiss students with German students' characteristics is 453.03. Although their performance decreases, it remains 13.99 points higher than what was actually observed in Germany. This difference is large in magnitude and statistically significant. Controlling for the language spoken at home widens

[19] Ammermüller (2007, 271) finds that "[a]n additional year of schooling adds [...] 38 points in Germany" for the overall student population.

the gap, leaving 17.05 points unexplained. By contrast, when the adjustment is performed including the *country of origin*, instead of the language indicator, the test score disparity narrows. Including both jointly, it amounts to 9.50 which is not statistically significant but relevant in magnitude.

Compared to the matching decompositions in Columns 6 to 9, the results are confirmed with higher point estimates for the unexplained part. Here, when conditioning on *country of origin* in Column 8, the differences are larger than in the BO decomposition. The gap amounts to 13.39 test score points when including both language and origin (Column 9). Differences to the BO decompositions might be explained by violations of linearity or validity out-of-support, since the BO decomposition procedure simply predicts values for empty covariate-cells.

Panel B presents the results for the reverse adjustment. The structure effect now measures the difference between children of immigrants in Switzerland and immigrants' children in Germany with adjusted characteristics. Interestingly, in Column 2, the adjusted gap is larger than the unconditional gap. This is because adjusting to the characteristics of immigrants' children in Switzerland causes even lower performance for students in Germany, 430.66 on average, than those of the actual second-generation immigrant population in Germany, demonstrating a negative composition effect. The structure effect increases from 26.44 to 30.43 when adding *German spoken at home* as an additional control in Column 3. Recall that this is more than half a school year in the sample of immigrants and almost a full year equivalent for the overall student population. In Column 4, adding the origin of the student's parents instead decreases the gap to 23.20 and controlling for both the gap is 27.25 – statistically significant and large in magnitude. Compared to the matching results in Columns 6-9, the effects show the same increasing pattern when the adjustment is performed on individual and family characteristics, and further increases when the language indicator is included. The specifications that include *country of origin* exhibit a smaller disparity in performance.

In sum, I find the gap that cannot be explained by differences in covariates to be positive in all, and large in magnitude and statistically significant in most specifications. The key finding is that the students in Switzerland outperform those in Germany at the mean. Of this disparity only a small part is attributable to differences in background characteristics. Noteworthy, including the language spoken at home increases the unexplained part in all specifications.

This suggests that Switzerland supports performance especially well for those who do not speak German at home. In the next section, I extend the analysis to the distribution of the reading test scores.

3.4.2 Distribution

The decompositions along the test score distribution are presented in Figures 3.2 and 3.3. Panel A presents the adjustments including the vector of *Other* background covariates and the *German spoken at home*, and Panel B adds the origin indicators. As in Figure 3.1, the left panel presents the reading test score densities and the right panel the quantile gaps. The adjusted quantile gaps are depicted by the solid lines and bootstrapped confidence intervals by the dotted lines (black for 95% and gray for 90% confidence intervals). As a reference, I add the unadjusted quantile gap (dashed line) from Figure 3.1.

— — — Figure 3.2 about here — — —

Starting with the adjustment of Swiss students to Germans characteristics, the density of immigrants' test scores in Switzerland does not align to the one of Germany. Accordingly, the gap remains roughly unchanged, as can be observed by the adjusted quantile gap in the right panel (solid line). This shows that, conditional on background characteristics, the large performance gap among low-performing children of immigrants remains and even widens slightly for the very well performing students. Including the origin of the parents aligns the adjusted Swiss students test score density more closely with the density of second-generation immigrants in Germany. Nevertheless, the gap for the very low-performing immigrants' children remains statistically significant and of considerable magnitude. The adjusted quantile gap is positive almost everywhere, though smaller in the medium percentiles and greater in the top percentiles than without taking the origin of the parents into account. Conditioning on origin increases the noise substantially rendering the gap only marginally significant at some parts of the distribution.

— — — Figure 3.3 about here — — —

The reverse adjustment is presented in Figure 3.3. In Panel A, the adjusted test score density among German students increases at a score of about 400 and decreases above 500 points. This projects an even lower performance for adjusted German students than the actual German students as discussed above. As shown in the right panel, the quantile gap increases in almost all quantiles. Including the *country of origin* indicators in Panel B, the density of German students adjusted to Swiss students' characteristics aligns more closely with the immigrants' children in Switzerland when the *country of origin* is included. Again, the very low-performing students are only found in Germany. Overall, the quantile gap appears to be positive in general, but is only marginally significantly different from zero.

Investigating the effects along the distributions of reading test scores, I find the structure effect to be largest among the very low-performing students. In all adjustments, the unexplained differences are substantial in magnitude for those that need the most support. The results confirm that performance is higher in Switzerland in all specifications and the gap remains mostly positive and marginally significant throughout the distribution when the *country of origin* is controlled for.

3.4.3 Subgroup analysis

"Which country supports the children of immigrants that face the most disadvantageous circumstances?" is one of the key questions for Western European policy makers, that are facing a growing number of children with migration background. "How do children descending from a specific origin compare to the results above?" and "does this disparity also exist between children of natives?" In Table 3.5, I address these questions by performing the decompositions separately on restricted samples of the student population, namely to: Those who have the most unfavorable background characteristics (Columns 1-3); children of Turkish immigrants, the larges overlapping immigrant population (Columns 4-6); and native students (Columns 7-9). Each first column presents the unconditional gap, the second the BO adjustment, and the third the matching adjustment.

— — — Table 3.5 about here — — —

In Columns 1-3, I restrict attention to second-generation immigrant students whose parents have either no education or only primary to lower secondary education (ISCED 0-2) and who do not speak German at home.[20] This procedure drastically reduces the sample to 296 observations – 57 in Germany and 212 in Switzerland – so results have to be interpreted cautiously. Unsurprisingly, the average test scores are lower in both countries (Column 1). Meanwhile, the unconditional gap increases sharply to 56.45 points, which is statistically significant and very large in magnitude. This subset of students performs better by more than a full year equivalent of schooling in Switzerland than in Germany. In Panel A Columns 2 and 3, when Swiss students are adjusted to German students' characteristics, the parametric and semi-parametric estimates of the structure effect are similar and substantial in magnitude, ranging from 30.15 to 36.78 test score points. Although parametric and matching estimates are still very large in magnitude, in the reverse decomposition (Panel B), they exhibit a greater dispersion ranging from 22.64 to 43.25 points. The gap is much larger in magnitude compared to the unconstrained sample of immigrants' children, suggesting that this subset is served much better in Switzerland than in Germany.

The parents' country of origin has been discussed as a potential explanation for some of the international variation in second-generation immigrant-native test score gaps. The children of Turkish descendants have attracted some attention since they represent a large enough population to be compared across countries (e.g., Dustmann, Frattini and Lanzara, 2012; Song, 2011). In Columns 4 to 6, I present the mean results for children of Turkish descent only, using all *Other* covariates and the *German spoken at home* indicator. Unconditionally, children of Turkish descent score 431.26 in Switzerland and 417.97 in Germany. While both scores are far below the international average of 500, they score 13.29 points higher in Switzerland than in Germany, although the difference is not statistically significant (probably due to the reduced sample size). In the Panel A decomposition, the unexplained gap amounts to 20.31 test score points, and in the reverse decomposition of Panel B to 34.22 test score points, which is statistically significant. The unexplained gap is large and consistently positive throughout the decompositions. The students – whose parents migrated from Turkey and who are subsequently born and raised within a German-speaking environment – in Switzerland substantially outperform those in Germany, especially after adjusting for background characteristics.

The economics of education literature has predominantly focused on educational integra-

[20] Applying these restrictions effectively constrains the books at home, since there are no observations with more than 200 books at home in Germany and only two in Switzerland.

tion, hence it is natural to ask what the adjustment mechanism returns when applied to natives' children. In Columns 7 to 9, I present the mean achievement gap for native students. First, it is important to note that the *parental education* category of *no primary education* is empty in Germany and nearly empty in Switzerland and consequently excluded from the specification. In Column 7, the average performance of Swiss children is better than that of German children, which would even increase some difference-in-differences measure of integration. Conditionally, this gap is reversed in both adjustments, meaning that the Swiss (German) students with German (Swiss) characteristics perform better (worse) than the observed Swiss (German) students, but only by a relatively small amount. Hence, natives' children also perform better in Switzerland, but notably less than their counterparts with a migration background.

3.4.4 Discussion

It is difficult to single out which institutional factors cause the large performance disparity, since educational institutions differ not only between but also within countries. I therefore end with a suggestive discussion of some features that might explain the large unexplained test score differences documented above. The decomposition results including school characteristics are presented in Table 3.6 (the descriptive statistics of the variables are presented in Table 3.1 above). Columns 3 and 7 of Table 3.4 are presented again in Columns 1 and 5 to serve as a benchmark.

— — — Table 3.6 about here — — —

Since the unexplained performance gap always increases when conditioning on language spoken at home, the Swiss system seems to be more capable in teaching its immigrants' children German reading literacy, especially to those who do not speak German at home. One possible explanation could be that – since all Swiss students do speak some dialect at home – the Swiss educational system has developed institutions that enhance second-language acquisition better than those in Germany. One way to assess whether there is a greater emphasis on learning German in the Swiss curriculum is to compare the amount of German language-lessons per week. On average, immigrant students in Switzerland report to visit more German classes as well as overall lessons per week than those in Germany (cf. Table 3.1). In Columns 2 and 7, I perform the decomposition using the *Other* background characteristics, the *German spoken at home*, the amount of German lessons, and the overall amount of lessons in school. It is important to note that there are a few missing values at the school level, hence the decompositions are based on a slightly reduced sample with average test scores of 456.22 in Switzerland and 442.49 in Germany. The unconditional gap amounts to 13.74 and essentially remains unaffected by the inclusion of the amount of lessons in all four decompositions. Notwithstanding, the amount of teaching is distinct from its quality which still might greatly influence the students' achievement. Interestingly, as shown in Table 3.1 above, the children in Germany visit twice as many out-of-school-time lessons in German.[21] It remains an open question whether the reading test score gap would even have been larger without the additional instruction time.

Early ability tracking is an intensively debated institution in both countries. The Swiss system tracks students between one and two years later in their school careers than the German system. This might cause the low performing students in Germany to fall behind their Swiss counterparts. On a distinct but related note, a result of early tracking could be segregation and placement of disadvantaged students in schools and/or classes. Similarly, Cattaneo and Wolter (2012) discuss that changes in the school composition – due to the increased number of children that speak German at home – can positively impact PISA performance of immigrant students with a migration background. Accordingly, if immigrant students are segregated they might have little reason or opportunity to learn German. Hence, the test score disparity may reflect the disparity in school compositions.

To address potential segregation, I calculate the proportions of second-generation immigrant students in school and define a categorical variable measured in 10% steps. On average in Germany, immigrants' children have 40% immigrants as peers in school, whereas the share is 34% in Switzerland (one should keep in mind that the overall share of immigrants is larger in Switzerland than in Germany). In Columns 3 and 8, when adjusting for the proportion of immigrants' children in school the unexplained gap reduces in all decompositions by a relatively small amount.

Another way to address the segregation is to control for the performance of the students' peers. Despite being highly endogenous, as explained by Manski (1993), it is still interesting

[21] The out-of-school-time lessons were assessed by the question: *How many hours do you typically spend per week attending out-of-school-time lessons in German (at school, at home or somewhere else)?* The answer categories were: 0, 0-2, 2-4, 4-6, 6 hours. To compare the averages, I take midpoint of the categories, 0 for the lowest, and 6 for the highest category.

to assess if peer performance can account for some of the unexplained test score gap. First, I calculate the average performance of second-generation immigrant students in school. In a second step, I additionally include native students.[22] The average reading test scores of the immigrant peers (and additional natives) is 470.27 (496.42) in Switzerland and 456.81 (476.54) in Germany. Including the performance of immigrants' children in the decomposition (Columns 4 and 9), the unexplained part of the gap reduces substantially. In addition, including the performance of natives' children (Columns 5 and 10), the disparity almost drops to 0, being insignificant in all specifications.

I do not want to over-emphasize these results due to the obvious endogeneity and the non-causal nature of the decomposition estimates. Yet, keeping in mind that the measure is problematic, the results point to a greater segregation and clustering of low-performing students in Germany. It seems promising to explore if the better performance is caused on the school level in greater detail. Furthermore, the finding highlights that if we seek to understand the second-generation immigrant students' achievement process, we have to find suitable comparison groups and a promising candidate being second-generation immigrant students in similar environments.

In sum, it appears that immigrants' children in German-speaking Switzerland exhibit greater reading (and math, cf. Appendix C.1) literacy than their counterparts in Germany. This difference prevails or even increases when adjusting for their personal, socio-economic, and educational characteristics. While the results are strong and unequivocal, their interpretation require caution, one must bear in mind that there might be other potential factors causing these differences in test scores which are not part of educational institutions, such as attitudes towards immigrants or integration efforts in general. In addition, it is of course possible that there is still some selection on unobservables. However, by decomposing the educational achievement gap conditionally on important background characteristics and using only those students that migrated from similar areas to similar countries that share the "same" language and migration history, it seems implausible that selection on unobservables alone causes these large differences in educational performance. Moreover, as the gap vanishes almost entirely by the inclusion of school system characteristics raises confidence in an explanation based on the educational systems.

[22] I use a categorical variable based on 25 test score point steps between 300 to 650. For the average peer test scores and the proportion of immigrants in school, results are robust to variations in the binning steps (results available upon request).

3.5 Conclusion

In this study I have proposed a new approach to investigate second-generation immigrants' learning processes by contrasting their performance across countries. Applying this reasoning, I have compared the immigrants' children in Germany with those in the German-speaking part of Switzerland to assure the students' comparability. Based on PISA 2009 survey results, I adjusted the test score distributions in Germany and German-speaking Switzerland for the composition of the second-generation immigrant population in the respective other country.

My results establish that the second-generation immigrants' reading literacy is substantially greater in Switzerland. This difference is large in magnitude, especially for low-performing students. The most crucial difference seems to be the language spoken at home. When it is different from German, it always increases the gap that cannot be explained by the students' background characteristics. These differences are robust to the inclusion of the parents' country of origin. Additionally, Switzerland seems to be particularly beneficial for unfavorably-endowed children of immigrants and the children of Turkish immigrants while being relatively less beneficial for children of native-born parents.

References

Algan, Yann, Christian Dustmann, Albrecht Glitz and Alan Manning. 2010. "The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom." *The Economic Journal* 120(542):F4–F30.

Ammermüller, Andreas. 2007. "PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany." *Empirical Economics* 33(2):263–287.

Ammermüller, Andreas, Hans Heijke and Ludger Wößmann. 2005. "Schooling quality in Eastern Europe: Educational production during transition." *Economics of Education Review* 24(5):579 – 599.

Belzil, Christian. and François Poinas. 2010. "Education and early career outcomes of second-generation immigrants in France." *Labour Economics* 17(1):101 – 110.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8(4):436–455.

Castles, Stephen. 1986. "The Guest-Worker in Western Europe - An Obituary." *International Migration Review* 20(4):761–778.

Cattaneo, Maria A. and Stefan C. Wolter. 2012. "Migration Policy Can Boost PISA Results: Findings from a Natural Experiment." *IZA Discussion Paper No. 6300.*

Dronkers, Jaap and Manon de Heus. 2010. Negative selectivity of Europe's guest-workers immigration? The educational achievement of children of immigrants compared with the educational achievement of native children in their origin countries. In *From Information to Knowledge; from Knowledge to Wisedom: Challenges and Changes facing Higher Education in the Digital Age*, ed. E. de Corte and J. Fenstad. London: Portland Press pp. 89 – 104.

Dustmann, Christian, Tommaso Frattini and Gianandrea Lanzara. 2012. "Educational achievement of second-generation immigrants: an international comparison." *Economic Policy* 27(69):143–185.

EDK. 2002. *Schweizerische Konferenz der kantonalen Erziehungsdirektoren.* Grundlegende Informationen zum Bildungswesen, Bern.

Entorf, Horst and Nicoleta Minoiu. 2005. "What a Difference Immigration Policy Makes: A Comparison of PISA Scores in Europe and Traditional Countries of Immigration." *German Economic Review* 6(3):355–376.

Fortin, Nicole M., Thomas Lemieux and Sergio Firpo. 2011. Decomposition Methods in Economics. In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card. Vol. 4, Part A Elsevier pp. 1 – 102.

Frölich, Markus 2007. "Propensity score matching without conditional independence assumption—with an application to the gender wage gap in the United Kingdom." *Econometrics Journal* 10(2):359–407.

Ganzeboom, Harry B. G., Paul M. De Graaf and Donald J. Treiman. 1992. "A standard international socio-economic index of occupational status." *Social Science Research* 21(1):1–56.

Häberlin, Urs, Christian Imdorf and Winfried Kronig. 2004. *Von der Schule in die Berufslehre: Untersuchungen zur Benachteiligung von ausländischen und von weiblichen Jugendlichen bei der Lehrstellensuche*. Berne: Haupt Verlag.

Hansen, Randall. 2003. "Migration to Europe since 1945: its history and its lessons." *Political Quarterly* 74(s1):25–38.

Heath, Anthony F., Catharine Rothon and Elina Kilpi. 2008. "The Second Generation in Western Europe: Education, Unemployment, and Occupational Attainment." *Annual Review of Sociology* 34:211–235.

KB. 2006. *Bildung in Deutschland*. Konsortium Bildungsberichterstattung. Bertelsmann Verlag GmbH & Co. KG, Bielefeld.

Levels, Mark, Jaap Dronkers and Gerbert Kraaykamp. 2008. "Immigrant Children's Educational Achievement in Western Countries: Origin, Destination, and Community Effects on Mathematical Performance." *American Sociological Review* 73(5):835–853.

Liebig, Thomas. 2004. Recruitment of foreign labour in Germany and Switzerland. In *Migration for Employment: Bilateral Agreements at a Crossroads*, ed. OECD, Integration Federal Office of Immigration and Emigration. OECD Publishing, Paris pp. 157–186.

Lüdemann, Elke and Guido Schwerdt. 2012. "Migration background and educational tracking." *Journal of Population Economics* 26(2):455–481.

Luthra, Renee R. 2010. *Assimilation in a New Context: Educational Attainment of the Immigrant Second Generation in Germany*. ISER Working Paper No. 2010-21.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60(3):531–542.

Mayda, Anna M. 2006. "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants." *Review of Economics and Statistics* 88(3):510–530.

Ñopo, Hugo. 2008. "Matching as a Tool to Decompose Wage Gaps." *Review of Economics and Statistics* 90(2):290–299.

Oaxaca, Ronald L. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3):693–709.

OECD. 2009. *PISA Data Analysis Manual (SPSS, second edition)*. Paris, OECD.

Ohinata, Asako and Jan C. van Ours. 2012. "Young Immigrant Children and their Educational Attainment." *Economic Letters* 116(3):288–290.

Van Ours, Jan C. and Jestus Veenman. 2003. "The educational attainment of second-generation immigrants in The Netherland." *Journal of Population Economics* 16(4):739–753.

Schmid, Carol. 1983. "Gastarbeiter in West Germany and Switzerland: an Assessment of Host Society-Immigrant Relations." *Population Research and Policy Review* 2:233–252.

Schütz, Gabriela, Heinrich W. Ursprung and Ludger Wößmann. 2008. "Education Policy and Equality of Opportunity." *Kyklos* 61(2):279–308.

Schneeweis, Nicole. 2011. "Educational institutions and the integration of migrants." *Journal of Population Economics* 24(4):1281–1308.

Schnepf, Sylke V. 2008. *Inequality of Learning Amongst Immigrant Children in Industrialised Countries*. IZA Discussion Paper No. 3337.

Segal, Carmit. 2012. "Working When No One Is Watching: Motivation, Test Scores, and Economic Success." *Management Science* 58(8):1438–1457.

Song, Steve. 2011. "Second-generation Turkish youth in Europe: Explaining the academic disadvantage in Austria, Germany, and Switzerland." *Economics of Education Review* 30(5):938–949.

Zimmermann, Klaus F. 1995. "Tackling the European Migration Problem." *Journal of Economic Perspectives* 9(2):45–62.

Tables and Figures

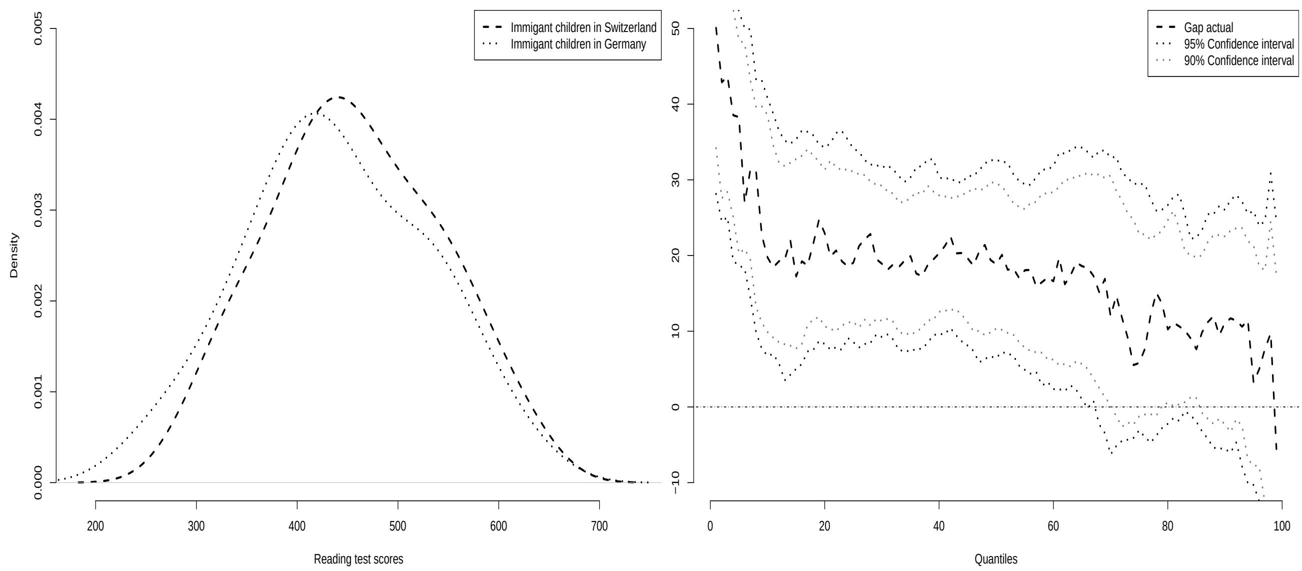
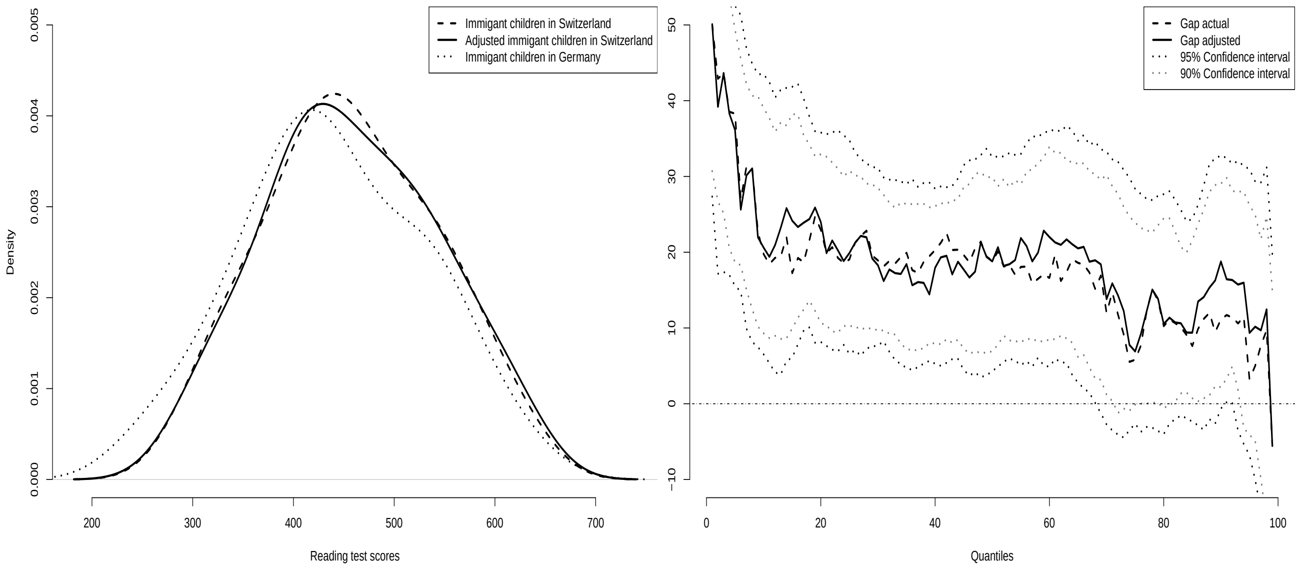
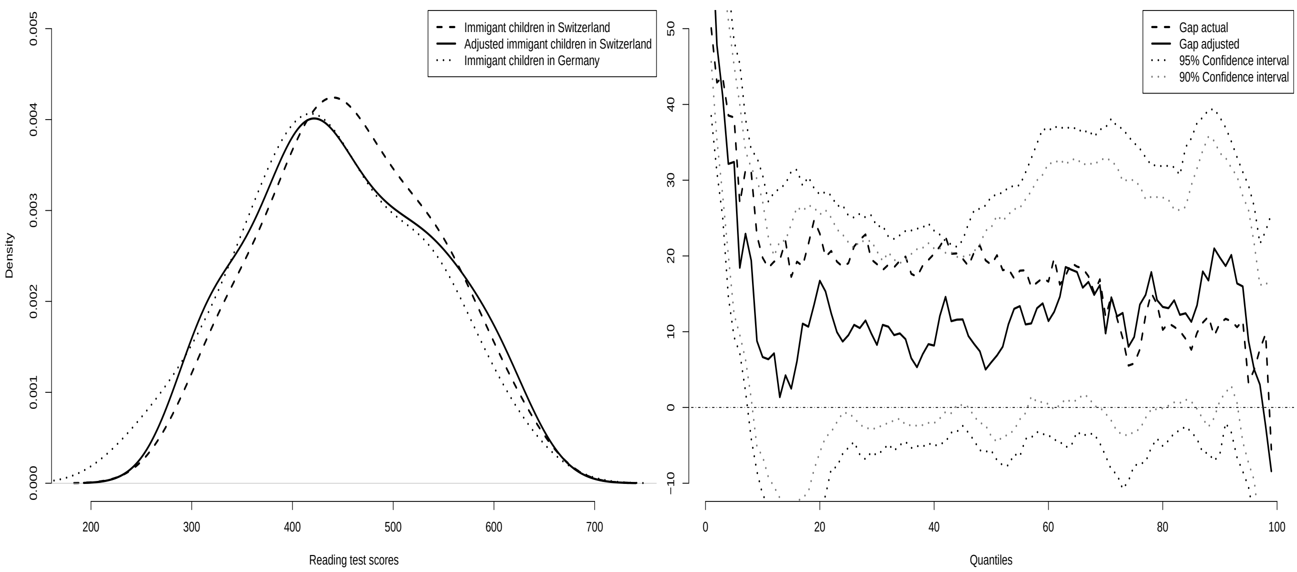


Figure 3.1: Second-generation immigrants kernel reading test score densities and unconditional quantile gap

Note: Left graph: Reading test score kernel densities; right graph: Quantile gap (dashed line) and bootstrapped confidence intervals (dotted lines: 90% gray; 95% black) based on 500 replications.
Source: PISA 2009, own calculations



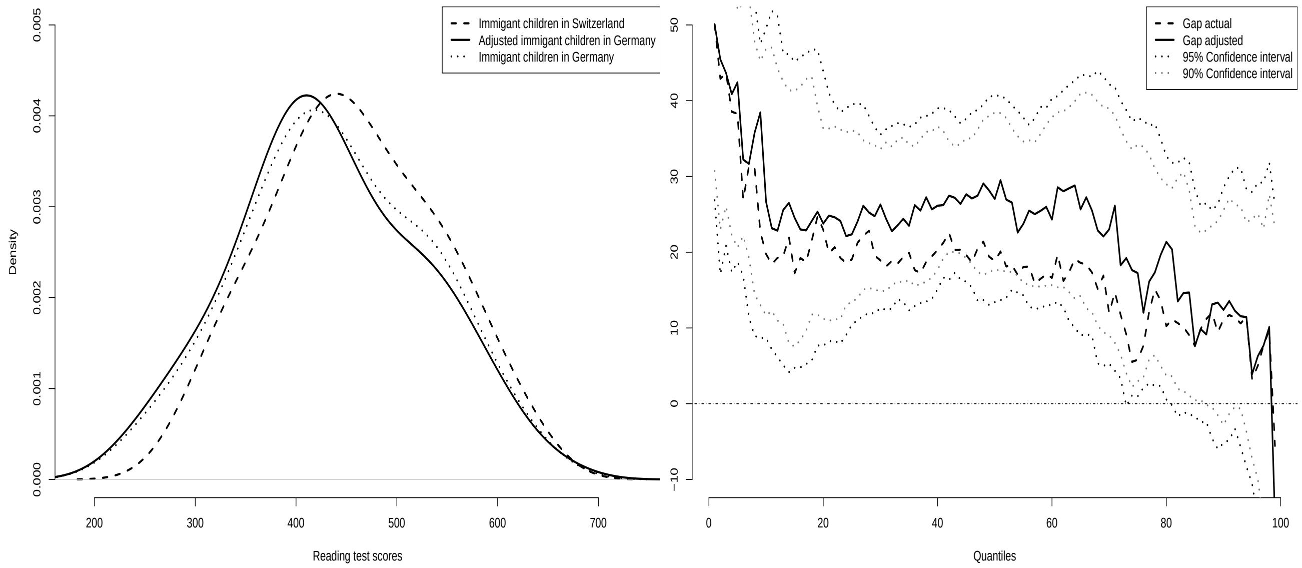
(A) Adjustment based on background characteristics and German spoken at home



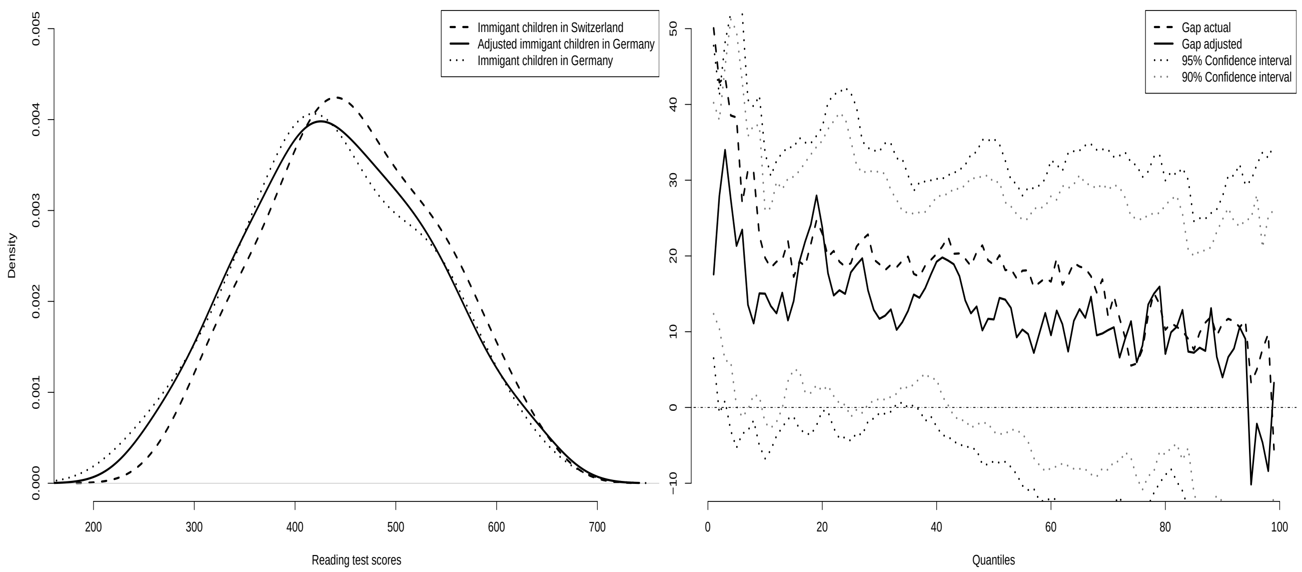
(B) Adjustment based on background characteristics, German spoken at home, and country of origin

Figure 3.2: Second-generation immigrants kernel reading test score densities and quantile gap, Swiss students adjusted to German characteristics

Note: Left graph: Reading test score kernel densities; right graph: Unconditional quantile gap (dashed line), adjusted quantile gap (solid line), and bootstrapped confidence intervals for the adjusted quantile gap (dotted lines) based on 500 replications; Panel A: adjustment is performed by propensity score matching including *gender*, *age in months*, *educational level of parents*, *highest occupation of parents*, *number of books at home*, and *German spoken at home.* Panel B: Additionally includes the parents' *country of origin.*
Source: PISA 2009, own calculations



(A) Adjustment based on background characteristics and German spoken at home



(B) Adjustment based on background characteristics, German spoken at home, and country of origin

Figure 3.3: Second-generation immigrants kernel reading test score densities and quantile gap, German students adjusted to Swiss characteristics

Note: Left graph: Reading test score kernel densities; right graph: Unconditional quantile gap (dashed line), adjusted quantile gap (solid line), and bootstrapped confidence intervals for the adjusted quantile gap (dotted lines) based on 500 replications; Panel A: adjustment is performed by propensity score matching including *gender*, *age in months*, *educational level of parents*, *highest occupation of parents*, *number of books at home*, and *German spoken at home*. Panel B: Additionally includes the parents' *country of origin*.
Source: PISA 2009, own calculations

Table 3.1: DESCRIPTIVE STATISTICS OF SCHOOL CHARACTERISTICS VISITED BY SECOND-GENERATION IMMIGRANT STUDENTS BY COUNTRY

Variables	Switzerland	Germany
Years visited preschool	2.79	2.67
	(0.44)	(0.61)
	[814]	[341]
Age at school entry	6.69	6.41
	(0.60)	(0.59)
	[778]	[335]
Grade at testing	8.80	8.97
	(0.57)	(0.73)
	[824]	[335]
Grade repetition	0.30	0.34
	[820]	[343]
Number of		
German lessons in school	4.34	4.10
	(1.10)	(0.99)
	[824]	[345]
all lessons in school	33.65	31.81
	(3.44)	(3.84)
	[809]	[343]
German lessons out-of-school	0.32	0.68
	(1.03)	(1.48)
	[593]	[219]
Proportion of migrants in school	0.34	0.40
Average reading test score		
immigrants' children in school	470.27	456.81
	(65.75)	(81.32)
+ natives in school	496.42	476.54
	(57.63)	(77.10)

Note: Switzerland refers to the German-speaking part only. School level variables imputed on school level first on immigrants' children responses and if still missing on those of natives as well. Standard deviations are given in round and number of observations if they differ from those in the main specification (CHE: 824; DEU: 356) in squared brackets. Grade repetition is an indicator variable that is one when the student repeated one or more grades and can be interpreted as percentage points. German lessons out-of-school was assessed categorically, the average is taken after redefinition by the midpoint of the categories that represent hours (the highest category was 6 and more, which is coded as 6). Proportion of second-generation migrants in school is calculated by the ratio of immigrants' children to immigrants' and native' children in school, and then binned into 10 percentage point steps. The average reading test scores are first calculated by using only the second-generation migrants in school and then additionally using the natives in school. These are then also binned into categories of 25 test score point steps from 300 to 650.
Source: PISA 2009, own calculations.

Table 3.2: Descriptive statistics of second-generation immigrant students by country

Variables	Switzerland	Germany
Reading literacy score	457.09	439.05
	(87.52)	(95.95)
Age (in months)	189.83	189.71
	(3.39)	(3.48)
Male	0.51	0.52
Education mother (ISCED)		
No education (0)	0.06	0.20
Primary (1,2)	0.42	0.21
Secondary (3,4)	0.29	0.42
Tertiary (5,6)	0.23	0.17
Education father (ISCED)		
No education (0)	0.03	0.16
Primary (1,2)	0.36	0.17
Secondary (3,4)	0.30	0.41
Tertiary (5,6)	0.31	0.26
Highest occupation (HISEI)	41.61	40.51
	(14.18)	(13.93)
Books at home		
0 - 10	0.29	0.31
11 - 25	0.26	0.21
26 - 100	0.29	0.28
101 - 200	0.09	0.12
201 - 500	0.05	0.05
More than 500	0.02	0.03
German spoken at home	0.19	0.33
Country of origin		
Southern Europe	0.14	0.06
Yugoslavia	0.48	0.07
Turkey	0.13	0.53
Another origin	0.16	0.29
Observations	824	356

Note: Switzerland refers to the German-speaking part only. All numbers are relative frequencies except for *reading literacy*, *age in months*, and *hisei*, for which standard deviations are given in round brackets. *Reading literacy* is the average of five plausible values provided by PISA. The parental education is summarized by four indicators according to the International Standard Classification of Education (ISCED), where *No education* (0) indicates that the parent has no formal education; *Primary* $(1, 2)$ captures primary and lower secondary education; *Secondary* $(3, 4)$ indicates upper secondary education and post-secondary (no tertiary education); and *Tertiary* $(5, 6)$ indicates any tertiary education. The highest occupational status of the parents ($HISEI$) as provided by PISA, measures highest occupational status of the parents where higher values correspond to occupations with high returns to education. Southern Europe includes Greece, Italy, Portugal, and Spain; Yugoslavia: Bosnia and Herzegovina, Croatia, FYR Montenegro and Serbia; Another country was a category in the PISA Questionnaire and includes parents originating from different areas.
Source: PISA 2009, own calculations.

Table 3.3: Second-generation immigrant-native students reading test score gaps

	Overall sample			*Selected sample*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Germany						
				without Aussiedler		
Immigrant	-58.32	-21.65	-7.91	-73.66	-32.31	-18.99
	(4.39)	(4.07)	(4.49)	(5.30)	(5.29)	(6.13)
Observations	3,791	3,791	3,791	3,632	3,632	3,632
Panel B: Switzerland						
				German-speaking		
Immigrant	-48.72	-17.86	-6.62	-54.06	-21.86	0.24
	(2.33)	(2.32)	(2.71)	(3.27)	(3.24)	(4.39)
Observations	8,292	8,292	8,292	5,249	5,249	5,249
				German-speaking without Western Europeans and Albanians		
Immigrant				-55.47	-21.72	-0.18
				(3.30)	(3.35)	(4.70)
Observations				5,189	5,189	5,189
Covariates						
Other	No	Yes	Yes	No	Yes	Yes
Language	No	No	Yes	No	No	Yes

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. Natives' children have both parents born in the country of testing. (1) and (4) report OLS regressions of reading test scores on an *Immigrant* indicator variable; (2) and (5) additionally include individual and family background characteristics (*Other*); (3) and (6) add the *German speaking at home* indicator, each for the respective sample selection. Robust standard errors are given in brackets.
Source: PISA 2009, own calculations.

Table 3.4: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics

	Actual	BO adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	457.09	453.03	456.09	445.78	448.54	455.56	458.22	451.26	452.44
Y_{DEU}	439.05								
Δ_X		4.06	1.00	11.31	8.55	1.53	-1.13	5.84	4.65
		(3.60)	(3.93)	(5.44)	(5.59)	(3.08)	(3.15)	(5.82)	(5.49)
Δ_S	18.05	13.99	17.05	6.74	9.50	16.52	19.18	12.21	13.39
	(5.78)	(6.05)	(5.74)	(6.79)	(7.13)	(5.71)	(5.76)	(7.24)	(7.09)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	457.09								
Y_{DEU}	439.05	430.66	426.67	433.90	429.85	438.00	433.84	444.75	443.97
Δ_X		-8.39	-12.38	-5.15	-9.20	-1.04	-5.21	5.70	4.92
		(4.75)	(5.47)	(9.09)	(9.62)	(3.76)	(4.27)	(6.98)	(7.03)
Δ_S	18.05	26.44	30.43	23.20	27.25	19.09	23.25	12.34	13.13
	(5.78)	(6.41)	(6.72)	(9.65)	(9.99)	(6.03)	(6.24)	(8.36)	(8.03)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,180	1,180	1,180	1,180	1,180	1,152	1,153	1,023	1,037
N_{CHE}	824	824	824	824	824	797	798	668	682
N_{DEU}	356	356	356	356	356	355	355	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin*, and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity scores are estimated by logit regressions on the same covariates as the in BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table 3.5: Mean gap decomposition between Swiss and German children reading test scores, adjusted to other countries characteristics

	Unfavorable background			Turkish			Natives		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	441.08	414.77	421.41	431.26	439.23	438.28	512.62	518.30	514.51
Y_{DEU}	384.63			417.97			513.52		
Δ_X		26.31	19.67		-7.97	-7.03		-5.68	-1.89
		(9.12)	(9.65)		(11.28)	(9.69)		(1.08)	(0.66)
Δ_S	56.45	30.15	36.78	13.29	21.26	20.31	-0.90	4.78	0.99
	(13.83)	(13.88)	(14.41)	(10.70)	(13.08)	(13.95)	(2.03)	(1.69)	(1.74)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	441.08			431.26			512.62		
Y_{DEU}	384.63	418.44	397.83	417.97	397.11	397.03	513.52	507.34	510.38
Δ_X		33.82	13.20		-20.87	-20.94		-6.18	-3.14
		(13.68)	(10.50)		(9.00)	(8.25)		(1.28)	(0.82)
Δ_S	56.45	22.64	43.25	13.29	34.15	34.22	-0.90	5.28	2.25
	(13.83)	(16.03)	(13.58)	(10.70)	(11.37)	(11.73)	(2.03)	(1.76)	(1.76)
Covariates									
*Other**	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
German	No	No	No	No	Yes	Yes	No	No	No
Observations									
N	269	269	196	296	296	287	7,600	7,600	7,594
N_{CHE}	212	212	140	109	109	105	4,362	4,362	4,359
N_{DEU}	57	57	56	187	187	182	3,238	3,238	3,235

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. Unconditional gap for immigrants' children with low parental background characteristics (1), Turkish descendants (4) and native students (both parents born in the country of testing) (7); *(2), (5) and (8) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based in (2) on *gender*, *age in month*, *educational level of parents* categories (cat.: 1 and 2), *highest occupation of the parents* and *number of books at home* (cat.: 1 to 3) and those that do not speak *German* at home; (5) uses all *Other* covariates of Table 3.4; (8) uses those of (5) without the primary education category for parental schooling; In (3), (6), and (9) adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity scores are estimated by logit regression on the same covariates as in the respective BO adjustments. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table 3.6: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics
School characteristics

	BO adjustment					Matching adjustment				
	(1)	(2)*	(3)	(4)	(5)	(6)	(7)*	(8)	(9)	(10)
Panel A: Swiss students adjusted to German students' characteristics										
Y_{CHE}	456.09	454.37	451.92	444.17	440.01	458.22	456.16	454.44	448.46	446.17
Y_{DEU}	439.05	442.49	439.05	439.05	439.05	439.05	442.49	439.05	439.05	439.05
Δ_X	1.00	2.72	5.17	12.92	17.08	-1.13	0.06	2.66	8.63	10.93
	(3.72)	(4.22)	(4.33)	(5.40)	(5.59)	(3.35)	(3.87)	(3.45)	(4.28)	(3.86)
Δ_S	17.05	11.01	12.88	5.13	0.97	19.18	13.67	15.39	9.42	7.12
	(5.93)	(6.02)	(5.59)	(4.05)	(4.66)	(5.74)	(6.04)	(5.38)	(4.16)	(4.92)
Panel B: German students adjusted to Swiss students' characteristics										
Y_{CHE}	457.09	456.22	457.09	457.09	457.09	457.09	456.22	457.09	457.09	457.09
Y_{DEU}	426.67	431.01	435.98	446.95	451.87	433.84	433.23	441.37	447.10	452.13
Δ_X	-12.38	-11.48	-3.07	7.90	12.82	-5.21	-9.25	2.33	8.05	13.08
	(5.82)	(5.80)	(5.58)	(5.55)	(5.81)	(4.48)	(5.44)	(4.91)	(5.02)	(5.41)
Δ_S	30.43	25.21	21.11	10.15	5.22	23.25	22.99	15.72	10.00	4.96
	(6.58)	(6.55)	(6.22)	(3.79)	(4.52)	(6.02)	(6.60)	(5.94)	(4.45)	(5.17)
Covariates										
Lessons	No	Yes	No	No	No	No	Yes	No	No	No
Prop. Mig.	No	No	Yes	No	No	No	No	Yes	No	No
Peer Migrants	No	No	No	Yes	No	No	No	No	Yes	No
+ Natives	No	No	No	No	Yes	No	No	No	No	Yes
Observations										
N	1,180	1,152	1,180	1,180	1,180	1,153	1,139	1,146	1,147	1,105
N_{CHE}	824	809	824	824	824	798	808	790	792	750
N_{DEU}	356	343	356	356	356	355	331	356	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. All specifications condition include *Other* covariates and *German spoken at home* indicator. (1) and (5) are the same as Columns (3) and (7) from Table 3.4. (2)* and (7)* include number of German lessons in school. In this decompositions the reference sample is reduced due to the inability to impute based on school level. The test score average before adjustment is 456.22 in Switzerland, 442.49 in Germany, leaving a raw gap of 13.74. (3) and (8) include the proportion of migrants in school. (4) and (9) add the average reading test score of immigrant students in school only, and (5) and (9) additionally include native students in school. (1) to (5) present BO adjustment. In (6) to (10), adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity scores are estimated by logit regression on the same covariates as the in BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Appendix A

Appendix: Chapter 1

A.1 Data

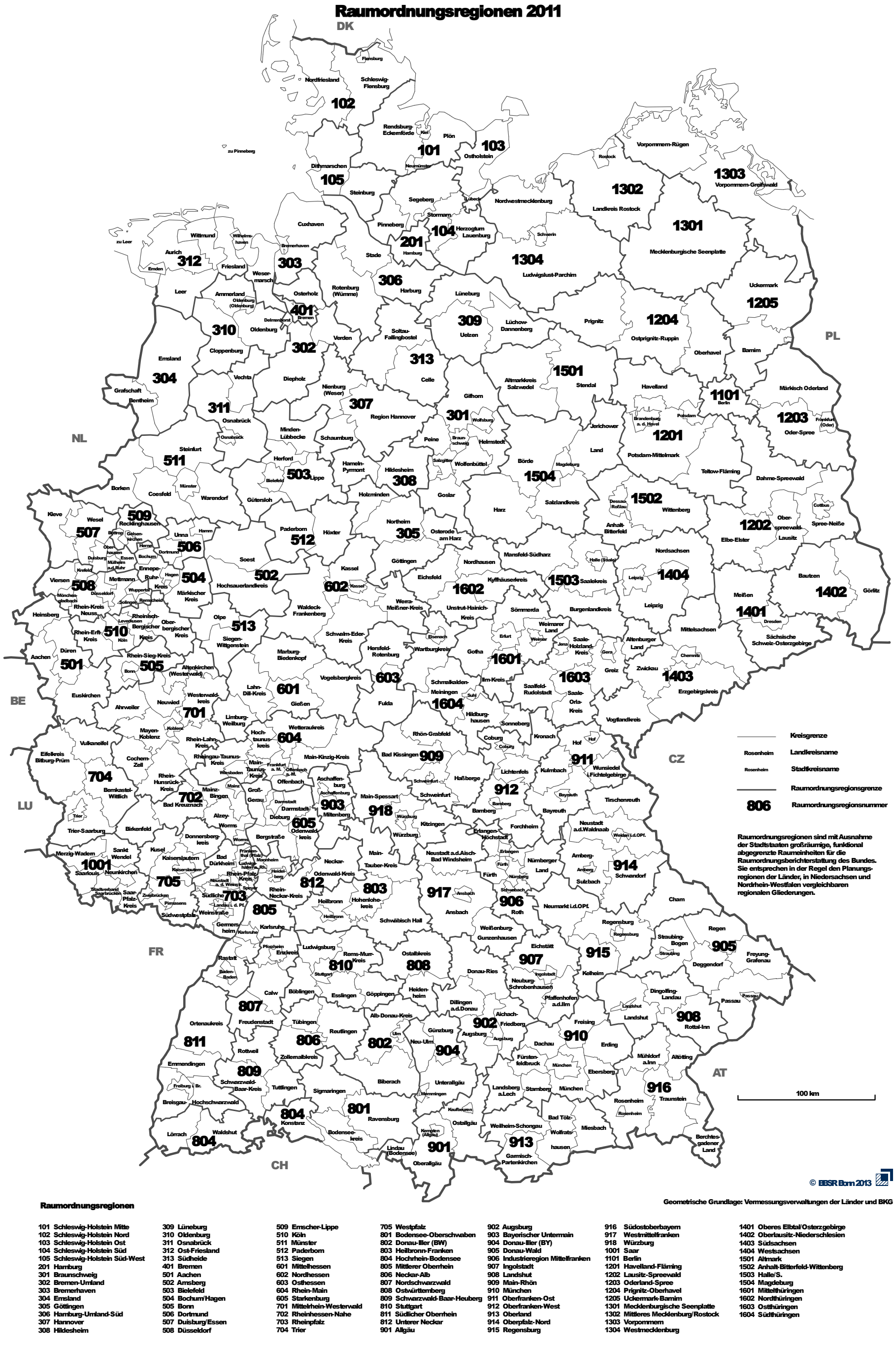


Figure A.1.1: LOCAL LABOR MARKETS, 96 RAUMORDNUNGREGIONEN [ROR]
Source: BBSR (2013)

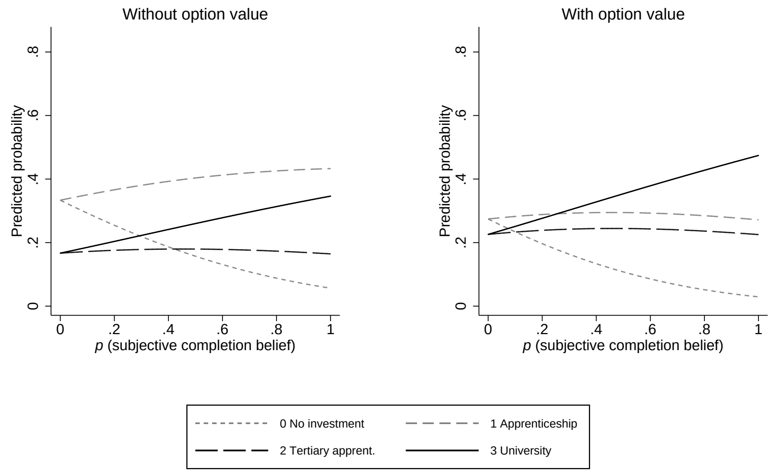


Figure A.1.2: Robustness: The role of the option value in a dynamic model of educational choice

Notes: Predicted probabilities constructed using estimates from Column (5) of Table 1.6 and evaluated at $x'\beta_j = 0$, $j = 1, 2, 3$. *Source*: SOEP 2000-2013, INKAR 2012, own calculations.

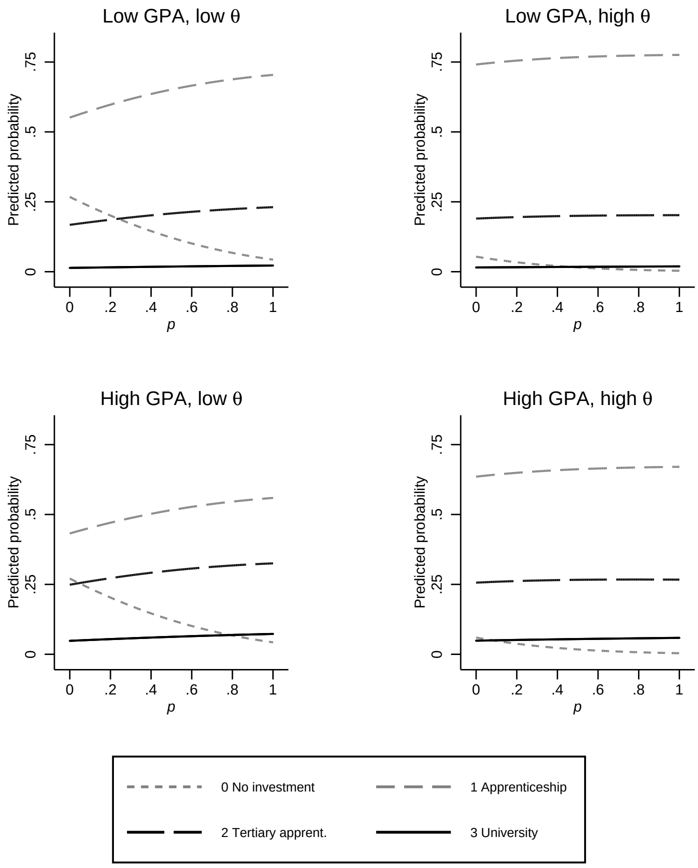


Figure A.1.3: Robustness: The role of academic ability and unobserved heterogeneity

Notes: Predicted probabilities constructed using estimates from Column (8) of Table 1.6, evaluated at $x'\beta_j = 0$, $j = 1, 2, 3$. High GPA = High $\theta = \Phi^{-1}(0.75)$, Low GPA = Low $\theta = \Phi^{-1}(0.25)$.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A.1.1: Descriptive statistics by subsample

	(A)	(B)	(C)	(D)	(F)
p	0.776	0.772	0.781	0.769	0.778
	(0.198)	(0.201)	(0.192)	(0.197)	(0.190)
GPA (std)	0.000	-0.018	0.012	0.041	0.077
	(1.000)	(1.019)	(1.009)	(1.024)	(1.016)
Rec: Lower Track (yes/no)	0.132	0.135	0.134	0.109	0.106
	(0.338)	(0.342)	(0.341)	(0.311)	(0.308)
Rec: Intermediate Track (yes/no)	0.259	0.272	0.270	0.247	0.249
	(0.438)	(0.445)	(0.444)	(0.431)	(0.433)
Rec: High school (yes/no)	0.402	0.347	0.360	0.384	0.400
	(0.490)	(0.476)	(0.480)	(0.487)	(0.490)
In high school (yes/no)	0.411	0.359	0.371	0.415	0.432
	(0.492)	(0.480)	(0.483)	(0.493)	(0.496)
Locus of control (std)	0.003	-0.038	-0.020	-0.053	-0.036
	(0.928)	(0.973)	(0.964)	(1.015)	(1.001)
Risk attitudes (std)	0.001	-0.085	-0.082	-0.153	-0.149
	(0.924)	(0.972)	(0.968)	(0.969)	(0.970)
Openness (std)	-0.000	-0.014	0.010	0.001	0.027
	(0.942)	(0.977)	(0.970)	(1.014)	(1.014)
Agreeableness (std)	0.002	-0.006	0.006	-0.011	0.006
	(0.944)	(0.984)	(0.980)	(0.995)	(1.002)
Extraversion (std)	-0.002	-0.010	0.009	-0.002	0.023
	(0.945)	(0.953)	(0.949)	(0.963)	(0.964)
Neuroticism (std)	-0.000	-0.015	-0.029	-0.027	-0.046
	(0.943)	(0.979)	(0.974)	(1.019)	(1.021)
Conscientiousness (std)	0.002	0.111	0.131	0.176	0.194
	(0.942)	(0.951)	(0.944)	(0.961)	(0.954)
Female (yes/no)	0.504	0.506	0.508	0.496	0.499
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Nr. siblings	1.570	1.639	1.630	1.633	1.622
	(1.316)	(1.339)	(1.323)	(1.335)	(1.322)
Second-generation migrant (yes/no)	0.623	0.739	0.731	0.843	0.840
	(0.485)	(0.439)	(0.443)	(0.364)	(0.367)
Parent college-educated (yes/no)	0.283	0.259	0.267	0.292	0.304
	(0.450)	(0.438)	(0.443)	(0.455)	(0.460)
Parent cur. unemployed (yes/no)	0.103	0.124	0.120	0.130	0.129
	(0.304)	(0.329)	(0.325)	(0.336)	(0.335)
Log. net household income	10.377	10.341	10.368	10.374	10.391
	(1.834)	(1.800)	(1.773)	(1.822)	(1.804)
Cyclical youth unemployment (in Ror)	0.074	0.211	0.211	0.267	0.260
	(1.026)	(1.034)	(1.035)	(1.083)	(1.082)
Nr. of apprenticeship positions (in Ror)	98.791	97.920	97.886	97.447	97.362
	(5.279)	(4.988)	(5.039)	(5.034)	(5.067)
Nr. of students (in Ror)	24.095	22.952	23.018	22.395	22.572
	(14.223)	(13.594)	(13.511)	(13.492)	(13.517)
Nr. of high school graduates (in Ror)	26.775	25.137	25.188	24.491	24.582
	(6.673)	(5.514)	(5.520)	(5.155)	(5.158)
Nr. of Universities (in Ror)	10.585	10.333	10.321	10.208	10.225
	(9.938)	(9.850)	(9.849)	(9.651)	(9.637)
N	3,610	2,116	1,919	1,372	1,255

Note: Table presents sample means and standard deviations in brackets in total and by subsample considered in Table 1.3.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A.1.2: Variable definitions

Variables	Description	Age	Missings
Core variables	*Missing values in the core variables are dropped from the estimation sample.*		
p	Subjective completion belief is a elicited measure, it ranges from 0 to 1, in 0.1 steps.	17	58
GPA (std)	Average of German and Math grades, standardized over the sample population, as a robustness check we additionally standardize within educational track (cf. Table A.2.3).	17	59
Educational outcomes:	From the longitudinal information we assess whether the student has started/completed a respective educational track.	17-31	-
$d \in \{0, 1, 2, 3\}$	Aspiration/Start/Complete, disaggregated by the tracks: drop out, apprenticeship, tertiary apprenticeship (high school and apprenticeship), and university (includes all higher learning institutions).		
$d_1 \in \{0, 1, 2\}$	First stage in structural model: drop out, apprenticeship, and high school.		
$d_2 \in \{0, 1\}$	Second stage in structural model: tertiary apprenticeship and university.		
Start apprenticeship	Not used in the analysis, all individuals that stared before are dropped from the estimation sample.	17	487
Still in school	Used in aspiration regressions, but dropped in the investment/completion analysis.	17	1,073
Academic variables			
Recommendations:	To visit a secondary-school track teachers evaluate the students (age the age of 10), the base category is no recommendation, three indicators for Lowest Track (yes/no), Intermediate Track (yes/no), and High school (yes/no)	17	249
In high school (yes/no)	An indicator whether the student is currently in high school when answering the youth questionnaire.	17	105
Personality variables	*We standardize the personality variables to mean 0 and standard deviation 1.*		
Locus of control (std)	First principal component of 10 questions, of which two are reversed.	17	459
Risk attitudes (std)	Assessed by a single question, ranging from 1-10.	17	306
Openness (std)	First principal component of 3 questions.	17	381
Agreeableness (std)	First principal component of 3 questions, of which one is reversed.	17	375
Extraversion (std)	First principal component of 3 questions, of which one is reversed.	17	378
Neuroticism (std)	First principal component of 3 questions, of which one is reversed.	17	378
Conscientiousness (std)	First principal component of 3 questions, of which one is reversed.	17	381
Individual and family characteristics	*Parental information, based on parents' questionnaires, are merged with the children's information.*		
Female (yes/no)	An indicator whether the individual is female.	17	
Nr. of siblings	Count of the number of siblings.	17	179
Second-generation migrant (yes/no)	An indicator whether the individual's parents are born in a foreign country, if information is missing recoded as second-generation migrant.	17	2,029
Parent college-educated (yes/no)	An indicator whether the individual has at least one college educated parent.	17	13

Parent cur. unemployed (yes/no)	An indicator whether the individual has at least one currently unemployed parent.	17	152
Log. net household income	Log of household pre-governmental income imputed by SOEP (0 income is treated as missing)	17	65
Fixed effects			
Year	Year of answering youth questionnaire, which is roughly identical to year of birth	17	
Region	Five regions based on federal states which are the level of educational-jurisdiction, cf. footnote 24 and Table A.2.3	17	109
Regional labor market information	*Information from INKAR 2012/Statistical agency, merged onto the students residence with 17 and lagged by one year. Some are twice assessed for the estimation of the structural model, based on residence with 17 to avoid endogeneity due to moving (there are no missings as the location is always known at 179.*		
Cyclical youth unemployment	Cyclical component of local youth unemployment, extracted by HP-filter.	16/18	
Nr. of apprenticeship positions	Number of apprenticeship positions by all potential apprentices times 100.	16/18	
Nr. of students	Number of students enrolled in higher learning institutions by all residents in the age group times 1000.	16/18	
Nr. of high school graduates	Number of students with a high school degree in the region over all school-leavers times 1000.	16/18	
Nr. of universities	Count of higher learning institutions in the Ror, due to minimal variation over time we keep it constant.	16	

Note: Table presents variable descriptions and missing values for the baseline sample. All available individuals add up to 4,192, which then reduce to 3,610. The remaining missings are conditional on the estimation sample. All variables besides core variables are included in the estimation along with missing value indicators. More information on the regional indicators can be found under `http://www.inkar.de`

A.2 Robustness

Table A.2.1: Determinants of subjective completion beliefs, fractional response regressions

	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.030	0.029	0.030
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.023	0.024	0.024	0.025
	(0.012)	(0.012)	(0.012)	(0.013)
Rec: Intermediate Track (yes/no)	0.057	0.054	0.049	0.051
	(0.009)	(0.009)	(0.009)	(0.010)
Rec: High school (yes/no)	0.045	0.040	0.033	0.036
	(0.009)	(0.009)	(0.009)	(0.010)
In high school (yes/no)	0.004	0.002	-0.005	-0.006
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.022	0.020	0.019
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.007	0.005	0.005
		(0.004)	(0.004)	(0.004)
Openness (std)		0.004	0.004	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.008	0.008	0.008
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.015	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		0.000	0.002	0.000
		(0.004)	(0.004)	(0.004)
Conscientiousness (std)		0.030	0.033	0.033
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.011	-0.012
			(0.007)	(0.007)
Nr. siblings			-0.002	-0.003
			(0.002)	(0.002)
Second-generation migrant (yes/no)			-0.023	-0.009
			(0.007)	(0.013)
Parent college-educated (yes/no)			0.007	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			-0.002	0.002
			(0.012)	(0.012)
Log. net household income			0.006	0.006
			(0.002)	(0.002)
N	3’610	3’610	3’610	3’610
$\bar{p}$	0.776	0.776	0.776	0.776
$SD(p)$	0.198	0.198	0.198	0.198
R^2_n	0.054	0.114	0.122	0.129
Academic	+	+	+	+
F(pval)	193.883 (0.000)	130.487 (0.000)	109.471 (0.000)	107.115 (0.000)
Personality	-	+	+	+
F(pval)		194.414(0.000)	201.546(0.000)	199.798(0.000)
Family Background	-	-	+	+
F(pval)			31.445(0.000)	19.705(0.012)
Labor market + FE	-	-	-	+
F(pval)				29.572(0.162)

Note: The Table presents Bernoulli pseudo-maximum likelihood with probit conditional expectation function, as proposed by Papke and Wooldridge (1996, 2008). We report marginal effects and robust standard errors in round brackets, our goodness of fit measure is a nonlinear R-squared measure and is calculated as the squared correlation coefficient between the estimated conditional expectation and the observed subjective beliefs: $R^2_n = corr(\hat{p}, p)^2$, where $\hat{p} = \Phi(x'\hat{\beta})$ due to the probit specification. The regressions of subjective beliefs are presented on varying sets of covariates, in (1) only on academic, (2) adds personality, (3) family background and individual characteristics, and (4) local labor market characteristics, region and time fixed effects (coefficients not presented). We present the unconditional mean $\bar{p}$ and standard deviation $SD(p)$ of the dependent variable.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A.2.2: Robustness: dichotomizing subjective beliefs ($p \geq 0.70$)

	probit				bivariate probit			
					$\rho = .05$	$\rho = .1$	$\rho = .2$	$\rho = .3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.380	0.321	0.280	0.274	0.188	0.103	-0.067	-0.236
	(0.066)	(0.069)	(0.069)	(0.070)	(0.070)	(0.070)	(0.069)	(0.067)
	[0.068]	[0.055]	[0.047]	[0.045]				
R^2_n	0.027	0.040	0.048	0.059				
$R^2_n(p)$	0.041	0.049	0.055	0.065				
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.793$, $SD(p) = 0.406$								
(B) Actual investment								
p	0.490	0.453	0.439	0.429	0.343	0.257	0.087	-0.084
	(0.107)	(0.111)	(0.114)	(0.120)	(0.119)	(0.119)	(0.117)	(0.115)
	[0.045]	[0.039]	[0.035]	[0.028]				
R^2_n	0.086	0.100	0.122	0.184				
$R^2_n(p)$	0.110	0.120	0.140	0.198				
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.789$, $SD(p) = 0.408$								
(C) Actual investment, conditional on intentions								
p	0.471	0.453	0.443	0.437	0.351	0.264	0.092	-0.080
	(0.119)	(0.122)	(0.125)	(0.131)	(0.131)	(0.130)	(0.129)	(0.126)
	[0.040]	[0.037]	[0.033]	[0.025]				
R^2_n	0.083	0.095	0.115	0.205				
$R^2_n(p)$	0.105	0.114	0.132	0.220				
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.805$, $SD(p) = 0.396$								
(D) Actual completion								
p	0.231	0.225	0.209	0.196	0.109	0.022	-0.153	-0.328
	(0.089)	(0.091)	(0.093)	(0.094)	(0.094)	(0.094)	(0.093)	(0.092)
	[0.092]	[0.089]	[0.083]	[0.078]				
R^2_n	0.087	0.093	0.102	0.123				
$R^2_n(p)$	0.091	0.097	0.105	0.125				
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.794$, $SD(p) = 0.405$								
(E) Actual completion, conditional on intentions								
p	0.259	0.271	0.269	0.259	0.171	0.083	-0.094	-0.271
	(0.097)	(0.099)	(0.100)	(0.102)	(0.102)	(0.101)	(0.101)	(0.099)
	[0.103]	[0.108]	[0.107]	[0.103]				
R^2_n	0.092	0.098	0.108	0.127				
$R^2_n(p)$	0.097	0.103	0.112	0.131				
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.809$, $SD(p) = 0.393$								
Academic	-	+	+	+	+	+	+	+
Personality	-	-	+	+	+	+	+	+
Family Background	-	-	-	+	+	+	+	+
Labor market	-	-	-	+	+	+	+	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. In the bivariate probit regressions we restrict the correlation between the errors to be 0.05, 0.1, 0.2, 0.3. For each outcome in Panels (A) to (E), we present McFadden's pseudo-R^2 with and without p, and sample statistics for the varying subsamples. In the appendix we present analogous probit and bivariate probit regressions.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A.2.3: ROBUSTNESS: GPA STANDARDIZED WITHIN HIGH SCHOOL AND USING FEDERAL STATES FIXED EFFECTS

	GPA, by hs attendance				Federal states fe			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.921	0.810	0.727	0.717	0.928	0.817	0.733	0.722
	(0.142)	(0.146)	(0.148)	(0.150)	(0.142)	(0.146)	(0.148)	(0.150)
	[0.140]	[0.122]	[0.108]	[0.104]	[0.141]	[0.123]	[0.109]	[0.106]
R_n^2	0.029	0.040	0.048	0.059	0.027	0.038	0.046	0.058
$R_n^2(p)$	0.049	0.055	0.059	0.069	0.048	0.054	0.058	0.069
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.776$, $SD(p) = 0.198$								
(B) Actual investment								
p	0.997	0.918	0.911	0.870	1.033	0.959	0.947	0.888
	(0.223)	(0.228)	(0.241)	(0.250)	(0.224)	(0.229)	(0.242)	(0.252)
	[0.069]	[0.062]	[0.056]	[0.044]	[0.071]	[0.064]	[0.057]	[0.045]
R_n^2	0.087	0.100	0.122	0.183	0.089	0.102	0.124	0.180
$R_n^2(p)$	0.113	0.121	0.141	0.199	0.117	0.124	0.144	0.197
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.772$, $SD(p) = 0.201$								
(C) Actual investment, conditional on intentions								
p	0.901	0.849	0.840	0.730	0.924	0.877	0.862	0.729
	(0.256)	(0.262)	(0.272)	(0.276)	(0.255)	(0.263)	(0.273)	(0.279)
	[0.058]	[0.053]	[0.048]	[0.031]	[0.058]	[0.054]	[0.049]	[0.032]
R_n^2	0.085	0.095	0.114	0.205	0.089	0.098	0.118	0.204
$R_n^2(p)$	0.104	0.110	0.129	0.214	0.108	0.115	0.133	0.213
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.781$, $SD(p) = 0.192$								
(D) Actual completion								
p	0.434	0.410	0.351	0.331	0.408	0.378	0.333	0.312
	(0.181)	(0.184)	(0.189)	(0.192)	(0.180)	(0.184)	(0.188)	(0.191)
	[0.172]	[0.162]	[0.139]	[0.131]	[0.162]	[0.150]	[0.132]	[0.124]
R_n^2	0.089	0.093	0.102	0.123	0.078	0.082	0.092	0.112
$R_n^2(p)$	0.092	0.096	0.104	0.124	0.081	0.085	0.093	0.113
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.769$, $SD(p) = 0.197$								
(E) Actual completion, conditional on intentions								
p	0.467	0.479	0.439	0.421	0.454	0.455	0.430	0.410
	(0.198)	(0.202)	(0.206)	(0.210)	(0.197)	(0.200)	(0.205)	(0.208)
	[0.185]	[0.189]	[0.174]	[0.167]	[0.180]	[0.180]	[0.170]	[0.162]
R_n^2	0.095	0.098	0.108	0.127	0.084	0.087	0.097	0.117
$R_n^2(p)$	0.099	0.101	0.110	0.129	0.088	0.090	0.099	0.119
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.778$, $SD(p) = 0.190$								
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A.2.4: Robustness: GPA standardized within federal states and including a fifth order polynominal in GPA

	GPA, std by federal states				GPA, polynomial			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.921	0.811	0.727	0.718	0.921	0.807	0.724	0.716
	(0.142)	(0.146)	(0.148)	(0.150)	(0.142)	(0.146)	(0.149)	(0.150)
	[0.140]	[0.122]	[0.108]	[0.105]	[0.140]	[0.121]	[0.108]	[0.104]
R_n^2	0.029	0.040	0.048	0.058	0.029	0.041	0.049	0.060
$R_n^2(p)$	0.049	0.055	0.059	0.069	0.049	0.056	0.060	0.071
Sample: $N = 3,610$, $\bar{d} = 0.908$, $\bar{p} = 0.776$, $SD(p) = 0.198$								
(B) Actual investment								
p	0.997	0.923	0.915	0.865	0.997	0.889	0.883	0.839
	(0.223)	(0.228)	(0.240)	(0.249)	(0.223)	(0.227)	(0.240)	(0.250)
	[0.069]	[0.062]	[0.056]	[0.044]	[0.069]	[0.057]	[0.051]	[0.040]
R_n^2	0.087	0.100	0.122	0.184	0.087	0.109	0.132	0.192
$R_n^2(p)$	0.113	0.121	0.141	0.199	0.113	0.128	0.149	0.206
Sample: $N = 2,116$, $\bar{d} = 0.956$, $\bar{p} = 0.772$, $SD(p) = 0.201$								
(C) Actual investment, conditional on intentions								
p	0.901	0.858	0.847	0.729	0.901	0.838	0.832	0.719
	(0.256)	(0.261)	(0.272)	(0.275)	(0.256)	(0.263)	(0.275)	(0.282)
	[0.058]	[0.054]	[0.049]	[0.031]	[0.058]	[0.050]	[0.045]	[0.030]
R_n^2	0.085	0.095	0.114	0.205	0.085	0.102	0.123	0.212
$R_n^2(p)$	0.104	0.110	0.129	0.214	0.104	0.117	0.137	0.221
Sample: $N = 1,919$, $\bar{d} = 0.961$, $\bar{p} = 0.781$, $SD(p) = 0.192$								
(D) Actual completion								
p	0.434	0.412	0.354	0.334	0.434	0.421	0.368	0.352
	(0.181)	(0.185)	(0.189)	(0.192)	(0.181)	(0.185)	(0.190)	(0.193)
	[0.172]	[0.163]	[0.140]	[0.133]	[0.172]	[0.167]	[0.146]	[0.140]
R_n^2	0.089	0.093	0.102	0.123	0.089	0.097	0.106	0.127
$R_n^2(p)$	0.092	0.096	0.104	0.124	0.092	0.099	0.108	0.129
Sample: $N = 1,372$, $\bar{d} = 0.544$, $\bar{p} = 0.769$, $SD(p) = 0.197$								
(E) Actual completion, conditional on intentions								
p	0.467	0.479	0.440	0.423	0.467	0.498	0.463	0.445
	(0.198)	(0.202)	(0.206)	(0.210)	(0.198)	(0.203)	(0.207)	(0.211)
	[0.185]	[0.190]	[0.174]	[0.167]	[0.185]	[0.197]	[0.183]	[0.176]
R_n^2	0.095	0.098	0.108	0.127	0.095	0.102	0.111	0.131
$R_n^2(p)$	0.099	0.101	0.110	0.129	0.099	0.105	0.114	0.133
Sample: $N = 1,244$, $\bar{d} = 0.547$, $\bar{p} = 0.778$, $SD(p) = 0.190$								
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table A.2.5: Robustness: separate regressions by high school attendance

	Not in high school				In high school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intention-to-invest								
p	0.988	0.880	0.802	0.801	0.655	0.446	0.424	0.466
	(0.160)	(0.164)	(0.167)	(0.168)	(0.303)	(0.311)	(0.324)	(0.335)
	[0.178]	[0.158]	[0.142]	[0.139]	[0.071]	[0.048]	[0.043]	[0.042]
R_n^2	0.018	0.030	0.042	0.052	0.030	0.042	0.059	0.088
$R_n^2(p)$	0.045	0.051	0.058	0.068	0.036	0.044	0.062	0.091
N	2'125	2'125	2'125	2'125	1'485	1'476	1'476	1'473
(B) Actual investment								
p	1.018	0.993	1.002	0.992	1.122	0.557	-0.013	-1.082
	(0.243)	(0.249)	(0.261)	(0.277)	(0.622)	(0.605)	(0.726)	(0.966)
	[0.097]	[0.092]	[0.086]	[0.068]	[0.036]	[0.010]	[0.000]	[0.000]
R_n^2	0.067	0.080	0.103	0.183	0.078	0.187	0.328	0.464
$R_n^2(p)$	0.097	0.106	0.128	0.204	0.099	0.191	0.328	0.469
N	1356	1356	1356	1355	584	582	582	577
(C) Actual investment, conditional on intentions								
p	0.950	0.979	0.992	0.934	0.710	0.088	-0.634	-3.554
	(0.281)	(0.285)	(0.291)	(0.311)	(0.653)	(0.651)	(0.823)	(1.551)
	[0.080]	[0.080]	[0.074]	[0.044]	[0.024]	[0.002]	[-0.002]	[0.000]
R_n^2	0.081	0.090	0.115	0.249	0.071	0.183	0.330	0.505
$R_n^2(p)$	0.103	0.113	0.136	0.264	0.078	0.183	0.333	0.538
N	1207	1207	1207	1206	546	544	544	539
(D) Actual completion								
p	0.534	0.501	0.420	0.367	0.247	0.193	0.213	0.230
	(0.212)	(0.215)	(0.219)	(0.225)	(0.347)	(0.359)	(0.380)	(0.391)
	[0.203]	[0.190]	[0.159]	[0.139]	[0.097]	[0.076]	[0.084]	[0.088]
R_n^2	0.055	0.062	0.071	0.103	0.115	0.120	0.141	0.170
$R_n^2(p)$	0.061	0.067	0.075	0.105	0.116	0.120	0.141	0.171
N	802	802	802	801	570	570	568	564
(E) Actual completion, conditional on intentions								
p	0.490	0.500	0.441	0.401	0.460	0.411	0.414	0.468
	(0.238)	(0.241)	(0.246)	(0.252)	(0.362)	(0.374)	(0.394)	(0.406)
	[0.183]	[0.187]	[0.165]	[0.150]	[0.180]	[0.161]	[0.162]	[0.179]
R_n^2	0.063	0.067	0.076	0.107	0.110	0.114	0.132	0.163
$R_n^2(p)$	0.068	0.072	0.080	0.109	0.112	0.115	0.134	0.165
N	709	709	709	708	535	535	533	529
Academic	-	+	+	+	-	+	+	+
Personality	-	-	+	+	-	-	+	+
Family	-	-	-	+	-	-	-	+
Labor market	-	-	-	+	-	-	-	+

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. *Source:* SOEP 2000-2013, INKAR 2012, own calculations.

Appendix B

Appendix: Chapter 2

B.1 Derivation of the probability function of the dynamic hurdle model

The probability of a zero in the DH model equals the probability of a zero in a Poisson model with rate λ_0, further visit follow a Poisson distribution with λ_1, which depends on the remaining time in the quarter $T-t$. The density can be written

$$\Pr(Y=k|\lambda_0,\lambda_1)=\int_0^T \exp(-\lambda_0 t)\lambda_0\exp(-\lambda_1(T-t))\frac{(\lambda_1(T-t))^{k-1}}{(k-1)!}dt.$$

If $\lambda_0=\lambda_1=\lambda$, it reduces to the standard Poisson:

$$\begin{aligned}\Pr(Y=k|\lambda)&=\int_0^T \exp(-\lambda t)\lambda\exp(-\lambda(T-t))\frac{(\lambda(T-t))^{k-1}}{(k-1)!}dt\\ &=\exp(-\lambda T)\frac{\lambda^k}{k!}\int_0^T k(T-t)^{k-1}dt\\ &=\exp(-\lambda T)\frac{\lambda^k}{k!}\left(-(T-t)^k\Big|_0^T\right)\\ &=\exp(-\lambda T)\frac{(\lambda T)^k}{k!}\end{aligned}$$

For example, if $\lambda_0<\lambda_1$ and $k=1$, the probability is

$$\begin{aligned}\Pr(Y=1|\lambda_0,\lambda_1)&=\int_0^T \exp(-\lambda_0 t)\exp(-\lambda_1(T-t))dt\\ &=\lambda_0\exp(-\lambda_1 T)\int_0^T \exp((\lambda_1-\lambda_0)t)dt\\ &=\lambda_0\exp(-\lambda_1 T)\frac{\exp((\lambda_1-\lambda_0)T)-1}{(\lambda_1-\lambda_0)}\\ &=\frac{\lambda_0}{\lambda_1-\lambda_0}[\exp(-\lambda_0 T)-\exp(-\lambda_1 T)].\end{aligned}\tag{B.1}$$

In general, for $k=1,2,3,\ldots$, we have

$$\begin{aligned}\Pr(Y=k|\lambda_0,\lambda_1)&=\int_0^T \exp(-\lambda_0 t)\lambda_0\exp(-\lambda_1(T-t))\frac{(\lambda_1(T-t))^{k-1}}{(k-1)!}dt\\ &=\lambda_0\lambda_1^{k-1}\exp(-\lambda_1 T)\int_0^T \exp((\lambda_1-\lambda_0)t)\frac{(T-t)^{k-1}}{(k-1)!}dt,\end{aligned}$$

integrating by parts, yields

$$\int_0^T \exp((\lambda_1-\lambda_0)t)\frac{(T-t))^{k-1}}{(k-1)!}dt = \int_0^T v'(t)u(t) = v(t)u(t) - \int_0^T v(t)u'(t)dt$$

$$u'(t) = (k-1)\frac{(T-t)^{k-2}}{(k-1)!} = \frac{(T-t)^{k-2}}{(k-2)!}$$

$$v(t) = \frac{\exp(\lambda_1-\lambda_0)^t}{\lambda_1-\lambda_0}$$

rewriting the term under the integral

$$\begin{aligned}
\Pr(Y=k|\lambda_0,\lambda_1) &= \lambda_0\lambda_1^{k-1}\exp(-\lambda_1 T) \\
&\quad \times \left(\left.\frac{\exp((\lambda_1-\lambda_0)t)(T-t)^{k-1}}{(\lambda_1-\lambda_0)(k-1)!}\right|_0^T + \int_0^T \frac{\exp((\lambda_1-\lambda_0)t)(T-t)^{k-2}}{(\lambda_1-\lambda_0)(k-2)!}dt \right) \\
&= -\frac{\lambda_0\lambda_1^{k-1}\exp(-\lambda_1 T)}{(\lambda_1-\lambda_0)(k-1)!}T^{k-1} \\
&\quad + \frac{\lambda_1}{\lambda_1-\lambda_0}\underbrace{\int_0^T \exp(-\lambda_0 t)\lambda_0\exp(-\lambda_1(T-t))\frac{\lambda_1^{k-2}(T-t)^{k-2}}{(k-2)!}dt}_{\Pr(Y=k-1|\lambda_0,\lambda_1)} \\
&= \underbrace{\frac{\lambda_1}{\lambda_1-\lambda_0}}_{\alpha}\Pr(Y=k-1|\lambda_0,\lambda_1) - \underbrace{\frac{\lambda_0\lambda_1^{k-1}\exp(-\lambda_1 T)}{(\lambda_1-\lambda_0)(k-1)!}T^{k-1}}_{c_k}
\end{aligned}$$

Solving the recursive equation $p_k = \alpha p_{k-1} + c_k$ and pluging in equation (1), we have

$$
\begin{aligned}
\Pr(Y = k|\lambda_0, \lambda_1) &= \alpha^{k-1} \Pr(k = 1|\lambda_0, \lambda_1) + \sum_{j=0}^{k-2} \alpha^j c_{k-j} \\
&= \frac{\lambda_1^{k-1}}{(\lambda_1 - \lambda_0)^{k-1}} \frac{\lambda_0}{\lambda_1 - \lambda_0} [\exp(-\lambda_0 T) - \exp(-\lambda_1 T)] \\
&\quad - \sum_{j=0}^{k-2} \frac{\lambda_1^j}{(\lambda_1 - \lambda_0)^j} \frac{\lambda_0 \lambda_1^{k-j-1} \exp(-\lambda_1 T)}{(\lambda_1 - \lambda_0)(k - j - 1)!} T^{k-j-1} \\
&= \frac{\lambda_0 \lambda_1^{k-1}}{(\lambda_1 - \lambda_0)^k} [\exp(-\lambda_0 T) - \exp(-\lambda_1 T)] \\
&\quad - \lambda_0 \lambda_1^{k-1} \exp(-\lambda_1 T) \sum_{j=0}^{k-2} \frac{1}{(\lambda_1 - \lambda_0)^{j+1}} \frac{T^{k-j-1}}{(k - j - 1)!} \\
&= \frac{\lambda_0 \lambda_1^{k-1}}{(\lambda_1 - \lambda_0)^k} [\exp(-\lambda_0 T) - \exp(-\lambda_1 T)] + \frac{\lambda_0 \lambda_1^{k-1}}{(\lambda_1 - \lambda_0)^k} \exp(-\lambda_1 T) \\
&\quad - \lambda_0 \lambda_1^{k-1} \exp(-\lambda_1 T) \sum_{j=0}^{k-1} \frac{1}{(\lambda_1 - \lambda_0)^{j+1}} \frac{T^{k-j-1}}{(k - j - 1)!} \\
&= \frac{\lambda_0 \lambda_1^{k-1}}{(\lambda_1 - \lambda_0)^k} \exp(-\lambda_0 T) - \lambda_0 \lambda_1^{k-1} \exp(-\lambda_1 T) \sum_{j=0}^{k-1} \frac{1}{(\lambda_1 - \lambda_0)^{j+1}} \frac{T^{k-j-1}}{(k - j - 1)!} \\
&= \frac{\lambda_0 \lambda_1^{k-1} \exp(-\lambda_0 T)}{(\lambda_1 - \lambda_0)^k} \left(1 - \exp(-(\lambda_1 - \lambda_0) T) \sum_{j=0}^{k-1} \frac{(\lambda_1 - \lambda_0)^{(k-1)-j}}{((k - 1) - j)!} T^{(k-1)-j} \right) \\
&= \frac{\lambda_0 \lambda_1^{k-1} \exp(-\lambda_0 T)}{(\lambda_1 - \lambda_0)^k} \left(1 - \sum_{j=0}^{k-1} \exp(-(\lambda_1 - \lambda_0) T) \frac{(\lambda_1 - \lambda_0)^j}{j!} T^j \right)
\end{aligned}
$$

Where from line 3 to 4, we use $\sum_{j=0}^{k-2} f_j(x) = \sum_{j=0}^{k-1} f_j(x) - f_{k-1}(x)$ and $k-j-1 = k-(k-1)-1 = 0$. In sum, the individual density of the dynamic hurdle model can be written as

$$
DHurdle(k, \lambda_0, \lambda_1, T) =
\begin{cases}
\exp(-\lambda_0 T) \frac{(\lambda_0 T)^k}{k!}, & \text{for } \lambda_0 = \lambda_1 \\
\exp(-\lambda_0 T)^{\mathbf{1}[k=0]} \times \left[\frac{\lambda_0 \lambda_1^{k-1} \exp(-\lambda_0 T)}{(\lambda_1 - \lambda_0)^k} \left(1 - \sum_{j=0}^{k-1} \exp(-(\lambda_1 - \lambda_0) T) \frac{(\lambda_1 - \lambda_0)^j}{j!} T^j \right) \right]^{\mathbf{1}[k>0]}, & \text{else.}
\end{cases}
$$

Dealing with mismatch

Consider the reporting period $(0 + r, T + r)$. Observations with $r = T$, follow dynamic hurdle process given above. Suppose the first sickness event occurred between 0 and r for which the visits are not reported, this occurs with probability given by $Pr(Y_A > 0) = 1 - \exp(-\lambda_0 r)$. Then the individual already payed the fee and has 0 monetary costs from visiting a doctor. Consequently, all visits between r and end of quarter T follow a Poisson distribution with $Poisson(s; \lambda_1(T - r))$. By contrast, if there was no previous visit $Pr(Y_A = 0) = \exp(-\lambda_0 r)$ there is simply a dynamic hurdle process with a reduced duration $DHurdle(s; \lambda_0, \lambda_1, T - r)$.

Additionally, the second period overlapping with the new quarter which follows a dynamic hurdle process $DHurdle(k-s;\lambda_0,\lambda_1,r)$. Since it is unknown when doctoral visits occurred, we sum over all possible combinations:

$$
\begin{aligned}
&DHurdle^{rep}(k,\lambda_0,\lambda_1,r)\\
&\quad= DHurdle(k;\lambda_0,\lambda_1,1)^{\mathbf{1}[r=T]} \times \left[\sum_{s=0}^{k} \Pr(Y_B=s)\Pr(Y_C=k-s)\right]^{\mathbf{1}[r\neq T]}\\
&\quad= DHurdle(k;\lambda_0,\lambda_1,T)^{\mathbf{1}[r=T]}\\
&\qquad\times\left[\sum_{s=0}^{k}[(1-\exp(-\lambda_0 r))\times Poisson(s;\lambda_1(T-r))+\exp(-\lambda_0 r)\times DHurdle(s;\lambda_0,\lambda_1,T-r)]\right.\\
&\qquad\left.\times DHurdle(k-s;\lambda_0,\lambda_1,r)\right]^{\mathbf{1}[r\neq T]}.
\end{aligned}
$$

Estimation

For our main estimation, we use longitudinal information and estimate the pooled model $l = i\times t$, the individual likelihood can be written

$$l_l(y_l;\lambda_{l,0},\lambda_{l,1},T) = DHurdle(y_l;\lambda_{l,0},\lambda_{l,1},T).$$

We replace the $DHurdle(\cdot)$ with the $DHurdle^{rep}(\cdot)$ for the reporting time mismatch.
Now we introduce log-normal unobserved heterogeneity, by

$$l_l(y_l;\lambda_{l,0},\lambda_{l,1},T) = \int DHurdle(y_l;\lambda_{l,0},\lambda_{l,1},T,\varepsilon_i)f(\varepsilon_i)\,\mathrm{d}\varepsilon_i$$

which we approximate by Gauss-Hermit Quadrature using m-quadrature points, exploiting the panel dimension with the assumption that the heterogeneity is constant across time. Due to numerical properties, we estimate the model in wide format by appropriately constraining the parameters to be equal across time periods. The log-likelihood can be written

$$\log L(y_{it};\sigma,\lambda_{it,0},\lambda_{it,1},T) = \sum_i \log\frac{1}{\sqrt{\pi}}\sum_m w^m \prod_t DHurdle(y_{it};\lambda_{it,0}\sqrt{2}a^m\sigma;\lambda_{it,1}\sqrt{2}a^m\sigma,T),$$

where a^m and w^m are quadrature abscissas and weights. We replace the $DHurdle(\cdot)$ with the $DHurdle^{rep}(\cdot)$ for the reporting time mismatch.

B.2 Monte Carlo Simulation

We illustrate the small sample properties of the dynamic hurdle model by Monte Carlo simulations.

Model 1

The data generating process is given by:

$$
\begin{aligned}
n &= 5,000; \\
x_{i1} &\sim Unifrom(0,2) \\
x_{i2} &= 1[x_{i1} + 0.2r_{i1} > 1], \text{ with } r_{i1} \sim N(0,1) \\
\lambda_{i0} &= \exp(-0.5 + 0.5x_{i1} - 0.2x_{i2}) \\
\lambda_{i1} &= \exp(0.1 + 0.8x_{i1} + 0.5x_{i2}) \\
t_i &= -\ln(r_{i2})/\lambda_{i0}, \text{ with } r_{i2} \sim Unifrom(0,1) \\
y_i &= \begin{cases} 0 & \text{if } t_i > 1, \\ 1 + r_{i4} & \text{otherwise, where } r_{i4} \sim Poisson(\lambda_{i1}(1-t_i)). \end{cases}
\end{aligned}
$$

In the Table B.2.1 we present the maximum likelihood estimation results of a poisson, a cloglog and zero truncated poisson (fixed hurdle), and our dynamic hurled regression, in that order.

Table B.2.1: Monte Carlo Simulation results: DGP 1 (1,000 replications)

	True		Poisson	Fixed Hurdle		Dynamic Hurdle	
	λ_0	λ_1	$\hat{\lambda}$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_0$	$\hat{\lambda}_1$
$Cons$	-0.5	0.1	-0.41	-0.50	0.04	-0.50	0.10
			(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
x_1	0.5	0.8	0.90	0.50	0.72	0.50	0.80
			(0.03)	(0.06)	(0.04)	(0.06)	(0.05)
x_2	-0.2	0.5	0.16	-0.20	0.38	-0.20	0.50
			(0.04)	(0.07)	(0.04)	(0.07)	(0.06)
ll			-10,640.75	-8,910.06		-8,800.90	

The results are reassuring even for a relatively small sample of 5,000 cross-sectional observations. The dynamic hurdle regression is performing well. For this basic example, the fixed hurdle also does relatively well at least qualitatively.

Model 2

The model is identical to Model 1, besides the log-normal unobserved heterogeneity which is generated by

$$\begin{aligned} \varepsilon_i &= \exp(-1.9) r_{i3}, \;\; \text{with } r_{i3} \sim N(0,1) \\ \lambda_{ij}^{\varepsilon} &= \lambda_{ij}\varepsilon_i, \qquad\qquad \text{for } j = 0,1 \end{aligned}$$

In Table B.2.2 we present the maximum likelihood estimation results of a poisson, a cloglog and zero truncated poisson (fixed hurdle), and our dynamic hurled regressions. Both zero truncated and the dynamic hurdle now allow for a log-normal unobserved heterogeneity, which is approximated by Gauss-Hermit quadrature.

Table B.2.2: MONTE CARLO SIMULATION RESULTS: DGP 2 (500 replications)

	True		Poisson	Fixed Hurdle		Dynamic Hurdle	
	λ_0	λ_1	$\hat{\lambda}$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_0$	$\hat{\lambda}_1$
$Cons$	-0.5	0.1	-0.40	-0.50	-0.02	-0.50	0.10
			(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
x_1	0.5	0.8	0.90	0.50	0.72	0.50	0.80
			(0.03)	(0.06)	(0.04)	(0.06)	(0.05)
x_2	-0.2	0.5	0.17	-0.20	0.39	-0.20	0.50
			(0.04)	(0.07)	(0.05)	(0.07)	(0.6)
$\ln(\sigma)$	-1.9				-1.00	-1.98	
					(0.05)	(0.39)	
ll			-10,866.09	-8,919.53		-8,898.51	

The dynamic hurdle model works well for this data generating process. Note, however, that the dynamic hurdle model did not converge in 14/500 replications.

Model 3

The model is identical to Model 1, besides that we vary the reporting time r_g such that $(r_g, T + r_g)$. Therefore, we generate two time periods (quarters) of 90 days. Within each time period $(0, T)$ and $(T, 2T)$ we draw the first and second event as above but allocating events to days in each quarter. We randomly select observations to one of two groups g, for which we set the reporting interval to $r_1 = 0$ and $r_2 = 15$, respectively.

In Table B.2.3 we present the maximum likelihood estimation results of a poisson, a cloglog and zero truncated poisson (fixed hurdle), and our dynamic hurled regressions. For the models besides the dynamic hurdle drop observations where $r_g = 15$.

Table B.2.3: Monte Carlo Simulation results: DGP 3 (500 replications)

	True		Poisson	Fixed Hurdle		Dynamic Hurdle	
	λ_0	λ_1	$\hat{\lambda}$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_0$	$\hat{\lambda}_1$
$Cons$	-0.5	0.1	-0.41	-0.50	0.05	-0.50	0.10
			(0.04)	(0.06)	(0.05)	(0.04)	(0.05)
x_1	0.5	0.8	0.90	0.50	0.72	0.50	0.80
			(0.05)	(0.08)	(0.05)	(0.06)	(0.05)
x_2	-0.2	0.5	0.15	-0.21	0.38	-0.21	0.49
			(0.05)	(0.10)	(0.06)	(0.07)	(0.06)
N			2,498	2,498	1,491	5,000	
ll			-5,303.06	-4,450.61		-8,825.88	

Similar to the previous data generating processes, the dynamic hurdle does very well estimating the parameters. Also the second advantage, of adjusting the the mismatch in terms of sample size comes into play here. Note, that in the simulation the r_g is independent of the covariates which is questionable in real applications, therefore, results are likely to be worse in practice for the other models.

Model 4

The model is identical to Model 1, besides both unobserved heterogeneity as in Model 2 and reporting time mismatch as in Model 3.

In Table B.2.4 we present the maximum likelihood estimation results of a poisson, a cloglog and zero truncated poisson (fixed hurdle), and our dynamic hurled regressions. Again number of observations are reduced for all models besides the dynamic hurdle model, and both the zero truncated poisson as well as our dynamic hurdle account for the unobserved heterogeneity.

Table B.2.4: Monte Carlo Simulation results: DGP 4 (50 replications)

	True		Poisson	Fixed Hurdle		Dynamic Hurdle	
	λ_0	λ_1	$\hat{\lambda}$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_0$	$\hat{\lambda}_1$
$Cons$	-0.5	0.1	-0.40	-0.49	-0.02	-0.49	0.10
			(0.04)	(0.06)	(0.06)	(0.04)	(0.05)
x_1	0.5	0.8	0.89	0.49	0.72	0.50	0.80
			(0.04)	(0.08)	(0.06)	(0.06)	(0.05)
x_2	-0.2	0.5	0.17	-0.20	0.39	-0.20	0.50
			(0.05)	(0.10)	(0.07)	(0.07)	(0.06)
$\ln(\sigma)$	-1.9				-1.00	-1.99	
					(0.07)	(0.40)	
N			2,498	2,498	1,491	5,000	
ll			-5,406.33	-4,449.53		-8,924.77	

Also for the most complicated data generating process the dynamic hurdle does very well even for the small sample size. Note, the dynamic hurdle did not converge in 3/100 cases.

Appendix C

Appendix: Chapter 3

C.1 Math literacy

In this Appendix, I present the analysis using math literacy scores as dependent variable (average of five Plausible Values), all other steps are equal to in the main specification. The results are highly significant economically and statistically. Qualitatively the math results confirm the reading results presented in the main text.

Table C.1.1: SECOND-GENERATION IMMIGRANT-NATIVE MATH TEST SCORE GAPS

	Overall sample			*Selected sample*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Germany						
				without Aussiedler		
Immigrant	-60.24	-23.09	-8.31	-76.25	-35.47	-20.23
	(4.37)	(4.24)	(4.83)	(5.21)	(5.47)	(6.51)
NObs	3,791	3,791	3,791	3,632	3,632	3,632
Panel B: Switzerland						
				German-speaking		
Immigrant	-62.47	-32.74	-22.21	-69.04	-36.23	-11.11
	(2.40)	(2.44)	(2.85)	(3.42)	(3.46)	(4.71)
NObs	8,292	8,292	8,292	5,249	5,249	5,249
				German-speaking without Western Europeans and Albanians		
Immigrant				-70.84	-36.73	-12.28
				(3.46)	(3.56)	(5.01)
NObs				5,189	5,189	5,189
Covariates						
Other	No	Yes	Yes	No	Yes	Yes
Language	No	No	Yes	No	No	Yes

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. Natives' children have both parents born in the country of testing. (1) and (4) report OLS regressions of reading test scores on an *Immigrant* dummy; (2) and (5) additionally include individual and family background characteristics (*Other*); (3) and (6) add the *German speaking at home* indicator, each for the respective sample selection. Robust standard errors are given in brackets. NObs stands for Number of Observations.
Source: PISA 2009, own calculations.

Table C.1.2: Mean gap decomposition between Swiss and German immigrants' children math test scores, adjusted to other countries characteristics

	Actual	BO adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	484.81	487.14	484.41	492.58	490.05	484.58	487.10	484.55	484.92
Y_{DEU}	452.15								
Δ_X		2.33	-0.40	7.77	5.24	0.22	-2.30	0.25	-0.12
		(3.78)	(4.02)	(5.71)	(5.70)	(3.12)	(3.19)	(5.94)	(5.61)
Δ_S	32.66	30.33	33.06	24.89	27.42	32.43	34.95	32.41	32.77
	(6.01)	(6.35)	(6.15)	(7.17)	(7.45)	(5.87)	(6.20)	(7.47)	(7.56)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	484.81								
Y_{DEU}	452.15	442.01	438.11	436.03	432.04	449.70	446.23	451.22	450.49
Δ_X		-10.14	-14.04	-16.12	-20.11	-2.45	-5.91	-0.93	-1.66
		(4.60)	(5.40)	(9.12)	(9.40)	(3.47)	(4.08)	(6.24)	(6.55)
Δ_S	32.66	42.80	46.70	48.78	52.76	35.11	38.57	33.58	34.32
	(6.01)	(6.51)	(6.93)	(9.57)	(9.83)	(5.92)	(6.44)	(8.06)	(8.10)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Number of Observations									
N	1,180	1,180	1,180	1,180	1,180	1,152	1,153	1,023	1,037
N_{CHE}	824	824	824	824	824	797	798	668	682
N_{DEU}	356	356	356	356	356	355	355	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin*; and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.1.3: Mean gap decomposition between Swiss and German children math test scores, adjusted to other countries characteristics

	Unfavorable background			Turkish			Natives		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	463.77	473.75	458.12	464.29	454.25	476.65	555.70	550.65	557.86
Y_{DEU}	414.99			434.60			529.25		
Δ_X		9.98	5.66		-10.04	-12.37		-5.05	-2.17
		(9.18)	(8.49)		(11.07)	(9.75)		(0.95)	(0.64)
Δ_S	48.79	38.81	43.13	29.68	39.72	42.05	26.45	31.50	28.61
	(11.88)	(13.81)	(13.29)	(10.89)	(12.97)	(13.47)	(2.08)	(1.78)	(1.93)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	463.77			464.29			555.70		
Y_{DEU}	414.99	432.59	423.00	434.60	414.24	417.87	529.25	522.82	526.04
Δ_X		17.60	8.01		-20.36	-16.73		-6.43	-3.21
		(13.10)	(8.75)		(7.82)	(6.78)		(1.20)	(0.83)
Δ_S	48.79	31.19	40.77	29.68	50.04	46.42	26.45	32.88	29.66
	(11.88)	(16.10)	(11.62)	(10.89)	(11.41)	(11.48)	(2.08)	(1.89)	(1.83)
Covariates									
*Other**	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
German	No	No	No	No	Yes	Yes	No	No	No
Number of Observations									
N	269	269	196	296	296	287	7,600	7,600	7,594
N_{CHE}	212	212	140	109	109	105	4,362	4,362	4,359
N_{DEU}	57	57	56	187	187	182	3,238	3,238	3,235

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. Unconditional gap for immigrants' children with low parental background characteristics (1), Turkish descendants (4) and native students (both parents born in the country of testing) (7); *(2), (5), and (8) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based in (2) on *gender*, *age in month*, *educational level of parents* categories (1) and (2), *highest occupation of the parents* and *number of books at home* (1) to (3) and those that do not speak *German* at home; (5) uses all *Other* covariates of Table 3.4; (8) uses those of (5) without the primary education category for parental schooling; (3), (6), and (9) Matching adjustment performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the respective BO adjustments. All standard errors given in brackets are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.1.4: Descriptive statistics of school characteristics visited by second-generation immigrant students by country

Variables	Switzerland	Germany
Number of		
Math lessons in school	4.62	4.22
	(1.10)	(1.06)
	[824]	[345]
all lessons in school	33.65	31.81
	(3.44)	(3.84)
	[809]	[343]
Math lessons out-of-school	0.55	0.79
	(1.25)	(1.49)
	[598]	[227]
Average math peer quality		
immigrants' children in school	497.03	466.57
	(71.31)	(79.41)
+ natives in school	530.95	487.85
	(62.58)	(79.13)

Note: Switzerland refers to the German-speaking part only. School level variables imputed on school level. Standard deviations are given in round and number of observations if they differ from those in the main specification (CHE: 824; DEU: 356) in squared brackets. Math lessons out-of-school was assessed categorically, the average is taken after redefinition by the midpoint of the categories that represent hours (the highest category was 6 and more, which is coded as 6). The average peer quality is first estimated by using only the second-generation migrants in school and then additionally using the natives in school. These are then binned into categories of 25 test score point steps from 300 to 650.
Source: PISA 2009, own calculations.

Table C.1.5: MEAN GAP DECOMPOSITION BETWEEN SWISS AND GERMAN IMMIGRANTS' CHILDREN MATH TEST SCORES, ADJUSTED TO OTHER COUNTRIES CHARACTERISTICS
SCHOOL CHARACTERISTICS

	BO adjustment					Matching adjustment				
	(1)	(2)*	(3)	(4)	(5)	(6)	(7)*	(8)	(9)	(10)
Panel A: Swiss students adjusted to German students' characteristics										
Y_{CHE}	485.21	484.58	479.53	457.83	447.55	487.10	482.59	482.57	465.45	461.93
Y_{DEU}	452.15	455.21	452.15	452.15	452.15	452.15	455.21	452.15	452.15	452.15
Δ_X	-0.40	-0.50	5.28	26.98	37.26	-2.30	1.50	2.24	19.35	22.88
	(3.92)	(4.54)	(4.48)	(5.55)	(5.99)	(3.32)	(3.94)	(3.81)	(4.22)	(4.44)
Δ_S	33.06	29.37	27.37	5.68	-4.60	34.95	27.37	30.42	13.31	9.78
	(6.35)	(6.60)	(5.53)	(4.05)	(5.07)	(5.96)	(5.95)	(5.82)	(4.62)	(5.26)
Panel B: German students adjusted to Swiss students' characteristics										
Y_{CHE}	484.81	484.08	484.81	484.81	484.81	484.81	484.08	484.81	484.81	484.81
Y_{DEU}	438.11	439.61	448.56	477.42	483.91	446.23	444.35	454.99	473.06	483.11
Δ_X	-14.04	-15.60	-3.59	25.27	31.76	-5.91	-10.86	2.84	20.92	30.96
	(5.55)	(5.62)	(5.40)	(5.77)	(5.96)	(4.18)	(5.17)	(4.69)	(5.71)	(5.72)
Δ_S	46.70	44.47	36.25	7.39	0.90	38.57	39.73	29.82	11.74	1.70
	(6.55)	(6.81)	(6.18)	(4.19)	(5.19)	(5.97)	(6.45)	(6.16)	(5.06)	(5.76)
Covariates										
Other	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lessons	No	Yes	No	No	No	No	Yes	No	No	No
Prop. Mig.	No	No	Yes	No	No	No	No	Yes	No	No
Peer quality										
Migrants	No	No	No	Yes	No	No	No	No	Yes	No
+ Natives	No	No	No	No	Yes	No	No	No	No	Yes
Observations										
N	1,180	1,152	1,180	1,180	1,180	1,153	1,137	1,146	1,143	1,105
N_{CHE}	824	809	824	824	824	798	808	790	788	750
N_{DEU}	356	343	356	356	356	355	329	356	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) and (5) are the same as Columns (3) and (7) from Table A.2. (2)* and (7)* include number of math lessons in school, in this decompositions the reference sample is reduced due to the inability to impute based on school level, the test score average before adjustment is 484.08 in Switzerland, 455.21 in Germany, and a raw gap of 28.87. (3) and (8) include the proportion of migrants in school. (4) and (9) the average peer reading test schore of migrants in school only, and (5) and (9) including additionally native children in school. (1) to (5) BO adjustment as described in Table 3. (6) to (10) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

C.2 Robustness checks

In this Appendix, I present the main Table 3 using reading literacy scores (average of five Plausible Values) as dependent variable. First, I include the French-, German- and Italian-speaking parts of Switzerland. Second, I present the main results for all immigrant students including Aussiedler, Albanians and Western Europeans. Third, I use sampling weights, Nearest Neighbor matching (1 and 5), and vary the bandwidth (0.095 and 0.105) of the main specification. Next, I use only the first plausible value, base the imputation mechanism on all observations within a country and use regressions (ordered probit and ordinary least squares) to predict the missing values.

Table C.2.1: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics: All areas

	Actual	BO adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	460.59	466.99	467.37	480.38	480.22	457.71	457.52	446.90	446.77
Y_{DEU}	439.05								
Δ_X		6.40	6.78	19.79	19.63	2.88	3.07	13.69	13.83
		(3.01)	(2.65)	(4.28)	(4.21)	(2.10)	(2.08)	(5.02)	(5.52)
Δ_S	21.55	15.15	14.76	1.76	1.92	18.67	18.47	7.86	7.72
	(5.75)	(5.88)	(5.27)	(5.95)	(5.91)	(5.25)	(5.37)	(6.45)	(7.08)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	460.59								
Y_{DEU}	439.05	434.74	435.76	437.09	437.43	440.14	440.61	449.06	449.12
Δ_X		-4.31	-3.29	-1.96	-1.62	1.09	1.56	10.01	10.07
		(4.81)	(4.79)	(7.84)	(7.95)	(3.34)	(3.26)	(6.20)	(5.83)
	21.55	25.86	24.84	23.51	23.17	20.46	19.99	11.54	11.48
	(5.75)	(6.18)	(6.17)	(8.31)	(8.53)	(5.45)	(5.84)	(7.18)	(7.43)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,962	1,962	1,962	1,962	1,962	1,940	1,950	1,742	1,753
N_{CHE}	1,606	1,606	1,606	1,606	1,606	1,584	1,594	1,387	1,398
N_{DEU}	356	356	356	356	356	356	356	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin*; and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.2.2: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics: All migrants

	Actual	BO adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	458.50	458.44	453.4	461.93	455.07	460.55	463.60	457.65	459.88
Y_{DEU}	454.39								
Δ_X		-0.06	-5.10	3.43	-0.61	-2.05	-5.10	0.85	-1.38
		(3.16)	(3.68)	(4.31)	(4.38)	(2.70)	(2.96)	(4.56)	(4.48)
Δ_S	4.11	4.17	9.21	0.68	4.72	6.16	9.21	3.26	5.49
	(5.21)	(5.20)	(5.09)	(5.62)	(5.25)	(5.35)	(5.21)	(5.94)	(5.85)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	458.50								
Y_{DEU}	454.39	441.16	434.22	443.42	438.40	447.88	442.68	455.32	452.22
Δ_X		-13.23	-20.17	-10.97	-16.35	-6.51	-11.71	0.93	-2.17
		(4.35)	(4.57)	(7.53)	(7.66)	(3.60)	(3.61)	(5.57)	(5.95)
Δ_S	4.11	17.34	24.28	15.09	20.46	10.62	15.82	3.18	6.28
	(5.21)	(5.54)	(5.34)	(8.19)	(8.34)	(5.37)	(5.38)	(6.50)	(6.75)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,399	1,399	1,399	1,399	1,399	1,373	1,375	1,261	1,274
N_{CHE}	884	884	884	884	884	859	863	747	760
N_{DEU}	515	515	515	515	515	514	512	514	514

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin* and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.2.3: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics: Sampling weights

	Actual	BO adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	456.60	462.12	458.81	475.44	472.38	455.84	458.58	451.29	452.45
Y_{DEU}	441.42								
Δ_X		5.52	2.21	18.84	15.78	0.77	-1.98	5.31	4.16
		(4.20)	(4.34)	(5.59)	(5.69)	(4.01)	(3.97)	(6.37)	(6.50)
Δ_S	15.18	9.66	12.97	-3.66	-0.60	14.41	17.16	9.87	11.02
	(5.83)	(6.28)	(6.31)	(7.07)	(7.10)	(5.74)	(5.59)	(7.73)	(7.31)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	456.60								
Y_{DEU}	441.42	432.64	429.34	436.79	433.73	437.90	433.66	444.84	443.73
Δ_X		-8.78	-12.08	-4.63	-7.69	-3.52	-7.77	3.42	2.30
		(5.06)	(5.26)	(8.87)	(8.92)	(4.50)	(4.66)	(6.75)	(6.93)
Δ_S	15.18	23.96	27.26	19.81	22.87	18.70	22.95	11.76	12.88
	(5.83)	(6.66)	(6.75)	(9.75)	(9.76)	(6.95)	(6.88)	(7.88)	(7.83)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,180	1,180	1,180	1,180	1,180	1,152	1,153	1,023	1,037
N_{CHE}	824	824	824	824	824	797	798	668	682
N_{DEU}	356	356	356	356	356	355	355	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin*; and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.2.4: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics: Nearest Neighbor matching

	Actual	Matching adjustment (1:1)				Matching adjustment (1:5)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	457.09	472.49	450.13	457.67	476.84	453.29	460.30	451.73	453.72
Y_{DEU}	439.05								
Δ_X		15.40	-6.96	0.58	19.75	3.80	-3.21	5.36	3.38
		(8.76)	(9.22)	(10.76)	(10.96)	(6.32)	(6.18)	(8.03)	(8.23)
Δ_S	18.05	2.64	25.01	17.47	-1.71	14.25	21.26	12.69	14.67
	(6.11)	(9.55)	(9.72)	(11.19)	(11.17)	(7.39)	(7.56)	(8.67)	(9.69)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	457.09								
Y_{DEU}	439.05	433.61	426.89	443.77	457.31	434.22	429.14	445.48	447.92
Δ_X		-5.44	-12.16	4.73	18.26	-4.82	-9.90	6.44	8.87
		(9.50)	(9.22)	(13.48)	(15.01)	(6.92)	(7.38)	(11.39)	(13.28)
Δ_S	18.05	23.49	30.20	13.32	-0.22	22.87	27.95	11.61	9.17
	(6.11)	(8.71)	(9.11)	(13.32)	(14.41)	(7.58)	(8.06)	(11.65)	(13.41)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,180	1,152	1,153	1,023	1,037	1,152	1,153	1,023	1,037
N_{CHE}	824	797	798	668	682	797	798	668	682
N_{DEU}	356	355	355	355	355	355	355	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) Unconditional gap; (2) to (9) Matching adjustment is performed by propensity score matching with Nearest Neighbor matching with 1 neighbor in (2) to (5) and 5 in (6) to (9), the propensity score is estimated by logit regression on *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home* in (2) and (5); (3) and (6) adds *German spoken at home*; (4) and (7) uses (2) and *country of origin*and (5) and (9) uses (2), (3), and (4); All standard errors given in brackets, are simulated with 500 bootstrap replications. **Note** that bootstrapping is not valid in the Nearest Neighbor approach. However for comparison reasons I stick to procedure.
Source: PISA 2009, own calculations.

Table C.2.5: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics: Bandwidth

	Actual	Matching adjustment (BW 0.095)				Matching adjustment (BW 0.105)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	457.09	455.43	458.20	451.05	452.38	455.69	458.24	451.47	452.51
Y_{DEU}	439.05								
Δ_X		1.66	-1.11	6.05	4.72	1.40	-1.15	5.62	4.58
		(3.33)	(3.27)	(5.95)	(5.93)	(3.11)	(3.13)	(5.39)	(5.68)
Δ_S	18.05	16.39	19.16	12.00	13.33	16.64	19.20	12.43	13.47
	(5.92)	(5.74)	(5.86)	(7.25)	(7.23)	(6.01)	(5.70)	(6.81)	(7.22)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	457.09								
Y_{DEU}	439.05	437.75	433.64	444.90	444.16	438.23	434.04	444.60	443.77
Δ_X		-1.29	-5.41	5.86	5.11	-0.81	-5.01	5.55	4.72
		(4.07)	(4.22)	(7.00)	(7.20)	(3.75)	(4.26)	(6.85)	(6.99)
Δ_S	18.05	19.34	23.46	12.19	12.93	18.86	23.05	12.50	13.33
	(5.92)	(6.02)	(6.08)	(7.87)	(8.09)	(6.41)	(6.44)	(7.97)	(7.86)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,180	1,152	1,153	1,023	1,037	1,152	1,153	1,023	1,037
N_{CHE}	824	797	798	668	682	797	798	668	682
N_{DEU}	356	355	355	355	355	355	355	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. (1) Unconditional gap; (2) to (9) Matching adjustment is performed by propensity score matching with Nearest Neighbor matching with 1 neighbor in (2) to (5) and 5 in (6) to (9), the propensity score is estimated by logit regression on *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home* in (2) and (5); (3) and (6) adds *German spoken at home*; (4) and (7) uses (2) and *country of origin*and (5) and (9) uses (2), (3), and (4); All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.2.6: Mean gap decomposition between Swiss and German immigrants' children reading test scores, adjusted to other countries characteristics: Plausible Value 1

	Actual	BO adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	455.31	451.66	454.62	444.28	446.92	454.15	456.73	449.20	450.42
Y_{DEU}	438.41								
Δ_X		3.65	0.69	11.03	8.39	1.16	-1.43	6.11	4.89
		(3.65)	(4.01)	(5.51)	(5.68)	(3.10)	(3.27)	(5.84)	(5.56)
Δ_S	16.89	13.24	16.21	5.86	8.51	15.73	18.32	10.79	12.01
	(5.89)	(6.23)	(5.85)	(6.90)	(7.31)	(5.81)	(5.92)	(7.39)	(7.31)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	455.31								
Y_{DEU}	438.41	428.32	424.42	435.46	431.50	435.75	431.47	444.31	442.88
Δ_X		-10.09	-13.99	-2.95	-6.91	-2.66	-6.94	5.89	4.47
		(4.93)	(5.61)	(9.29)	(9.98)	(3.88)	(4.42)	(7.05)	(7.22)
Δ_S	16.89	26.98	30.88	19.84	23.80	19.55	23.83	11.00	12.43
	(5.89)	(6.72)	(7.08)	(9.76)	(10.42)	(6.18)	(6.47)	(8.55)	(8.12)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,180	1,180	1,180	1,180	1,180	1,152	1,153	1,023	1,037
N_{CHE}	824	824	824	824	824	797	798	668	682
N_{DEU}	356	356	356	356	356	355	355	355	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. Instead of the main specification where the average of five Plausible Values is used, I present here the analysis based on only the first Plausible Value. (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin*; and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.2.7: MEAN GAP DECOMPOSITION BETWEEN SWISS AND GERMAN IMMIGRANTS' CHILDREN READING TEST SCORES, ADJUSTED TO OTHER COUNTRIES CHARACTERISTICS: IMPUTATION ON COUNTRY LEVEL

	Actual	Matching adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	457.09	452.03	454.94	444.60	447.27	454.42	456.92	449.82	450.04
Y_{DEU}	439.05								
Δ_X		5.06	2.15	12.49	9.82	2.67	0.17	7.27	7.06
		(3.81)	(4.04)	(5.54)	(5.69)	(3.17)	(3.34)	(6.18)	(5.97)
Δ_S	18.05	12.99	15.90	5.56	8.22	15.38	17.88	10.77	10.99
	(5.78)	(6.20)	(5.81)	(6.84)	(7.22)	(5.88)	(5.99)	(7.57)	(7.48)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	457.09								
Y_{DEU}	439.05	439.10	435.2	444.57	440.78	441.77	438.64	449.60	447.84
Δ_X		0.05	-3.85	5.52	1.73	2.73	-0.40	10.56	8.79
		(5.40)	(5.96)	(9.60)	(10.05)	(4.26)	(4.62)	(6.81)	(6.77)
Δ_S	18.05	18.00	21.90	12.53	16.32	15.32	18.45	7.49	9.26
	(5.78)	(6.74)	(7.02)	(9.93)	(10.16)	(6.18)	(6.50)	(8.02)	(7.65)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,180	1,180	1,180	1,180	1,180	1,144	1,132	997	970
N_{CHE}	824	824	824	824	824	790	777	642	614
N_{DEU}	356	356	356	356	356	354	355	355	356

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. Instead of using the children in school to impute missing values I use all children in the country. (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin*; and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

Table C.2.8: MEAN GAP DECOMPOSITION BETWEEN SWISS AND GERMAN IMMIGRANTS' CHILDREN READING TEST SCORES, ADJUSTED TO OTHER COUNTRIES CHARACTERISTICS: REGRESSION BASED IMPUTATION

	Actual	BO adjustment				Matching adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Swiss students adjusted to German students' characteristics									
Y_{CHE}	457.09	453.73	455.59	446.13	447.17	457.09	459.03	452.97	452.94
Y_{DEU}	439.05								
Δ_X		3.36	1.50	10.96	9.92	0.00	-1.93	4.13	4.15
		(3.79)	(3.96)	(5.58)	(5.61)	(3.23)	(3.42)	(5.83)	(5.66)
Δ_S	18.05	14.69	16.54	7.08	8.13	18.04	19.98	13.92	13.89
	(5.78)	(6.08)	(5.86)	(6.96)	(7.19)	(5.86)	(5.82)	(7.29)	(7.26)
Panel B: German students adjusted to Swiss students' characteristics									
Y_{CHE}	457.09								
Y_{DEU}	439.05	433.74	429.97	439.55	437.09	441.76	437.04	452.10	451.76
Δ_X		-5.31	-9.08	0.50	-1.96	2.72	-2.00	13.05	12.71
		(4.80)	(5.45)	(9.01)	(9.36)	(3.96)	(4.45)	(6.96)	(6.98)
Δ_S	18.05	23.36	27.13	17.55	20.00	15.33	20.05	5.00	5.34
	(5.78)	(6.50)	(6.45)	(9.22)	(9.57)	(6.12)	(6.34)	(8.00)	(7.84)
Covariates									
Other	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
German	No	No	Yes	No	Yes	No	Yes	No	Yes
Origin	No	No	No	Yes	Yes	No	No	Yes	Yes
Observations									
N	1,180	1,180	1,180	1,180	1,180	1,146	1,152	1,042	1,057
N_{CHE}	824	824	824	824	824	790	797	688	702
N_{DEU}	356	356	356	356	356	356	355	354	355

Note: Immigrants' children are born within the country of testing and have both parents born outside of the country. Instead of the main specification where the missing values are imputed based on median imputation within the school; I use ordered probit models based on all other variables in the model to predict first *number of books at home*, including the predicted number of books I again use an ordered probit regression iteratively for mothers' and then fathers' education, a probit regression for *German spoken at home* and am ordinary least squares regression for HISEI (rounding the predicted values to the next integer). (1) Unconditional gap; (2) to (5) BO adjustment: uses a twofold Blinder-Oaxaca decomposition based on (2) *gender*, *age in month*, *educational level of parents*, *highest occupation of the parents* and *number of books at home*; (3) adds *German spoken at home*; (4) uses (2) and *country of origin*; and (5) uses (2), (3), and (4); (6) to (9) Matching adjustment is performed by propensity score matching with Gaussian kernel and bandwidth 0.1, the propensity score is estimated by logit regression on the same covariates as in the BO adjustment. All standard errors given in brackets, are simulated with 500 bootstrap replications.
Source: PISA 2009, own calculations.

C.3 Sample selection and covariate balance

C.3.1 Missing values

First, I present the missing values, their imputations and the sample selection procedure. Next, I show the covariate balance before and after the matching and propensity scores over the common support.

Table C.3.1: Reading test scores by missing values

Variable	Switzerland			Germany		
	All	*Without missings*	*Only missings*	*All*	*Without missings*	*Only missings*
Reading test scores	457.09	473.28	424.60	439.05	471.91	415.93
	(87.52)	(86.74)	(79.85)	(95.95)	(88.47)	(94.28)
N	824	556	268	356	151	205

Note: Switzerland refers to the German-speaking part only. Reading test score is the average of five plausible values. If information of the student where missing in any of the variables *mother*'s or *father*'s education, *language spoken at home*, *number of books at home* or *highest occupational status of the parents*.
Source: PISA 2009, own calculations.

Table C.3.2: Missing value correlations matrix

Variables		Missing				
	Reading	*Mother*	*Father*	*German*	*Books*	*HISEI*
Reading test scores	1.000					
Missing values in						
Mother	-0.178	1.000				
Father	-0.138	0.491	1.000			
German	-0.167	-0.001	-0.008	1.000		
Books	-0.131	0.066	0.023	0.043	1.000	
HISEI	-0.147	0.155	0.224	0.082	0.105	1.000

Note: Switzerland refers to the German-speaking part only. Correlation matrix between missing value indicators and reading test scores.
Source: PISA 2009, own calculations.

Table C.3.3: Number of missing values and imputations

Variables	Migrants		Natives	
	CHE	DEU	CHE	DEU
Full sample				
Total	824	356	4,362	3,238
Mother education (ISCED)	50	109	176	143
Father education (ISCED)	67	89	173	262
German not spoken at home	206	77	105	97
Books at home	9	6	40	40
Highest occupation (HISEI)	26	39	36	115
Subsample: Unfavorable background				
Total	212	57		
Mother education (ISCED)	1	8		
Father education (ISCED)	1	6		
German not spoken at home	63	20		
Books at home	2	2		
Highest occupation (HISEI)	6	12		
Subsample: Turkish				
Total	109	187		
Mother education (ISCED)	9	64		
Father education (ISCED)	12	49		
German not spoken at home	29	39		
Books at home	1	2		
Highest occupation (HISEI)	6	19		

Note: Switzerland refers to the German-speaking part only. Western Europe includes Austria, France, Germany and Liechtenstein; Southern Europe includes Greece, Italy, Portugal, and Spain; Former Yugoslavia: Bosnia and Herzegovina, Croatia, FYR Montenegro and Serbia; *Aussiedler* includes Poland and former USSR and Another country represents a category in the PISA Questionnaire and includes parents originating from different areas. Imputed data based on school level median (including native students); *German* not spoken at home includes missing values.
Source: PISA 2009, own calculations.

Table C.3.4: Sample selection

Variables	Migrants		Natives	
	CHE	DEU	CHE	DEU
Observations (before selection)	1,697	515	6,598	3,276
Dropped due to				
Gender	1		1	
Parents no education (ISCED=0)			4	38
Non-German-speaking part	812		2,231	
Country of origin				
Albania	23			
Aussielder		159		
Western European	37			
Observations (after selection)	824	356	4,362	3,238

Note: Switzerland refers to the German-speaking part only. Western Europe includes Austria, France, Germany and Liechtenstein; *Aussiedler* includes immigrants from Poland and former USSR. Imputed data based on school level median (including native students); *German* not spoken at home includes missing values. *Source:* PISA 2009, own calculations.

C.3.2 Covariate balance

Table C.3.5: Covariate balance: Adjusting Swiss students to German students' characteristics, Panel A adjustments

Variables	Actual	Adjusted CHE				Actual
	DEU	(1)	(2)	(3)	(4)	CHE
Age (in months)	189.71	189.70	189.74	189.69	189.77	189.83
Male	0.52	0.53	0.53	0.55	0.54	0.51
Education mother (ISCED)						
No education (0)	0.20	0.15	0.15	0.18	0.18	0.06
Primary (1,2)	0.21	0.23	0.23	0.21	0.20	0.42
Secondary (3,4)	0.42	0.42	0.42	0.39	0.40	0.29
Tertiary (5,6)	0.17	0.20	0.20	0.22	0.22	0.23
Education father (ISCED)						
No education (0)	0.16	0.10	0.09	0.12	0.12	0.03
Primary (1,2)	0.17	0.19	0.19	0.16	0.16	0.36
Secondary (3,4)	0.41	0.43	0.44	0.42	0.44	0.30
Tertiary (5,6)	0.26	0.29	0.28	0.29	0.28	0.31
Highest occupation (HISEI)	40.51	41.49	41.49	42.01	42.04	41.61
Books at home						
0-10	0.31	0.30	0.29	0.31	0.30	0.29
11-25	0.21	0.22	0.22	0.22	0.22	0.26
26-100	0.28	0.28	0.29	0.29	0.30	0.29
101-200	0.12	0.11	0.11	0.11	0.11	0.09
201- 500	0.05	0.06	0.05	0.05	0.04	0.05
More than 500	0.03	0.03	0.03	0.03	0.03	0.02
German spoken at home	0.33	-	0.30	-	0.30	0.19
Country of origin						
Southern Europe	0.06	-	-	0.07	0.07	0.14
Yugoslavia	0.07	-	-	0.09	0.09	0.48
Turkey	0.53	-	-	0.45	0.44	0.13
Another origin	0.29	-	-	0.34	0.34	0.16
Observations	356					824

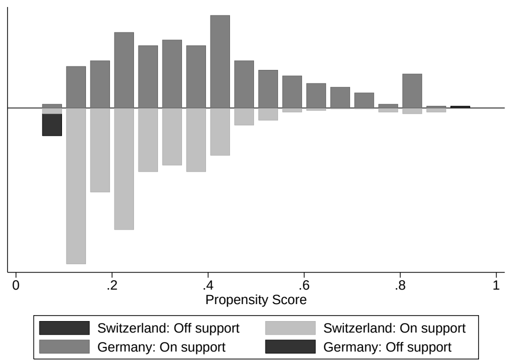
Note: Switzerland refers to the German-speaking part only. Besides the *age* in month and the *hisei* the variables can be interpreted as percentage points of the immigrants children population in each country. Columns (1) to (4) correspond to the adjustments of Table 3.4.
Source: PISA 2009, own calculations.

Table C.3.6: Covariate balance: Adjusting German students to Swiss students' characteristics, Panel B adjustments

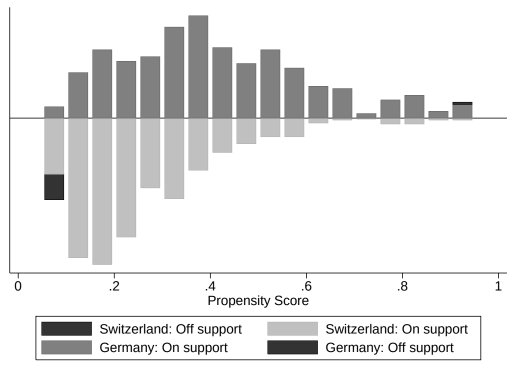
Variables	Actual	Adjusted DEU				Actual
	CHE	(1)	(2)	(3)	(4)	DEU
Age (in months)	189.83	189.76	189.87	189.88	189.94	189.71
Male	0.51	0.53	0.53	0.53	0.53	0.52
Education mother (ISCED)						
No education (0)	0.06	0.08	0.08	0.12	0.12	0.20
Primary (1,2)	0.42	0.31	0.31	0.29	0.29	0.21
Secondary (3,4)	0.29	0.34	0.34	0.30	0.30	0.42
Tertiary (5,6)	0.23	0.27	0.27	0.29	0.29	0.17
Education father (ISCED)						
No education (0)	0.03	0.04	0.04	0.06	0.06	0.16
Primary (1,2)	0.36	0.25	0.26	0.24	0.24	0.17
Secondary (3,4)	0.30	0.37	0.36	0.36	0.35	0.41
Tertiary (5,6)	0.31	0.34	0.33	0.34	0.34	0.26
Highest occupation (HISEI)	41.61	42.81	42.42	43.88	44.28	40.51
Books at home						
0-10	0.29	0.28	0.27	0.29	0.29	0.31
11-25	0.26	0.25	0.25	0.23	0.22	0.21
26-100	0.29	0.29	0.30	0.30	0.31	0.28
101-200	0.09	0.10	0.10	0.09	0.09	0.12
201- 500	0.05	0.05	0.05	0.05	0.05	0.05
More than 500	0.02	0.02	0.03	0.03	0.03	0.03
German spoken at home	0.19	-	0.23	-	0.26	0.33
Country of origin						
Southern Europe	0.14	-	-	0.14	0.15	0.06
Yugoslavia	0.48	-	-	0.28	0.28	0.07
Turkey	0.13	-	-	0.18	0.18	0.53
Another origin	0.16	-	-	0.24	0.25	0.29
Observations	824					356

Note: Switzerland refers to the German-speaking part only. Besides the *age* in month and the *hisei* the variables can be interpreted as percentage points of the immigrants children population in each country. Columns (1) to (4) correspond to the adjustments of Table 3.4.
Source: PISA 2009, own calculations.

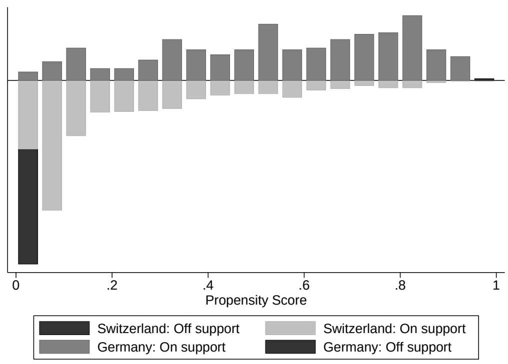
C.3.3 Common support



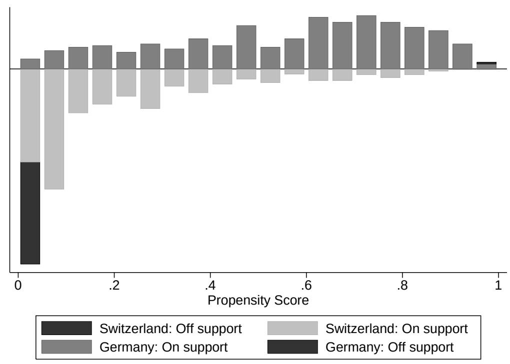
(1) Propensity score based on background characteristics



(2) Propensity score based on background characteristics and German spoken at home



(3) Propensity score based on background characteristics and country of origin



(4) Adjustment based on background characteristics, German spoken at home, and country of origin

Figure C.3.1: Common support graphs

Note: (1) Propensity scores are estimated on *gender*, *age in months*, *educational level of parents*, *highest occupation of parents*, and *number of books at home*; (2) Additionally uses *German spoken at home*; (3) uses parents' *country of origin* instead; (4) uses both *German spoken at home* and *country of origin.*
Source: PISA 2009, own calculations

Curriculum Vitae

Johannes Sebastian Matthias Nikolaus Kunz

Born July 19^{th} 1985
in Karlsruhe, Germany
German national

Education

2011 - 2016	Doctoral program at the *Zurich Graduate School of Economics*, University of Zurich, Switzerland
2014	Invited research stay at the University of Melbourne, Australia (Dept. of Economics)
2009 - 2011	Master of Science in *Quantitative Economics* at the University of Konstanz, Germany
2010	Exchange semester at the National University of Singapore, Singapore (Dept. of Economics) Scholarship: Baden-Württemberg-STIPENDIUM (Landesstiftung)
2006 - 2009	Bachelor of Arts in *Political and Administrative Science* at the University of Konstanz, Germany
2008	Exchange semester at the Universitat Pompeu Fabra, Spain (Dept. of Economics) Scholarship: Erasmus

Professional experience

2011 - 2016	Research and teaching assistant at the Department of Economics, University of Zurich, Switzerland
2009	Research assistant in the initiative *The Science of Social Stress and Conflict Resolution project: The Impact of Decisional Stress on the Escalation of Conflicts*, Chair of Prof. Dr. Gerald Schneider, University of Konstanz, Germany
2009	Internship at the YouGovPsychonomics AG in the *Department of Organisational Research & Consulting*, Cologne, Germany
2008	Internship at the Deutsche Gesellschaft für technische Zusammenarbeit GmbH (GTZ) in the project: *Strengthening of Institutional Development in the Road Sector* at the Ministry of Works and Transport, Windhoek, Namibia
2007 - 2008	Trainee at the business administration sector of the Siemens AG in the field *Industrial Solutions - Postal Automation*, Konstanz, Germany

March, 2016